PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/12, 1/21, 1/19, C07K 14/47, 16/18

A1

(11) International Publication Number: WO 95/20654

(43) International Publication Date: 3 August 1995 (03.08.95)

(21) International Application Number: PCT/GB95/00192

(22) International Filing Date: 31 January 1995 (31.01.95)

(30) Priority Data:

9401791.0	31 January 1994 (31.01.94)	GB
9418234.2	9 September 1994 (09.09.94)	GB

(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, MX, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ).

(71) Applicant *(for all designated States except US)*: THE INSTITUTE OF CANCER RESEARCH: ROYAL CANCER HOSPITAL [GB/GB]; 17A Onslow Garden, London SW7 3AL (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: WILLISON, Keith, Robert [GB/GB]; 79 Chesson Road, West Kensington, London W14 9QS (GB). KUBOTA, Hiroshi [JP/GB]; 2 Milton Road, Highgate, London N6 5QD (GB). ASHWORTH, Alan [GB/GB]; 88B Drayton Gardens, South Kensington, London SW10 9RF (GB).

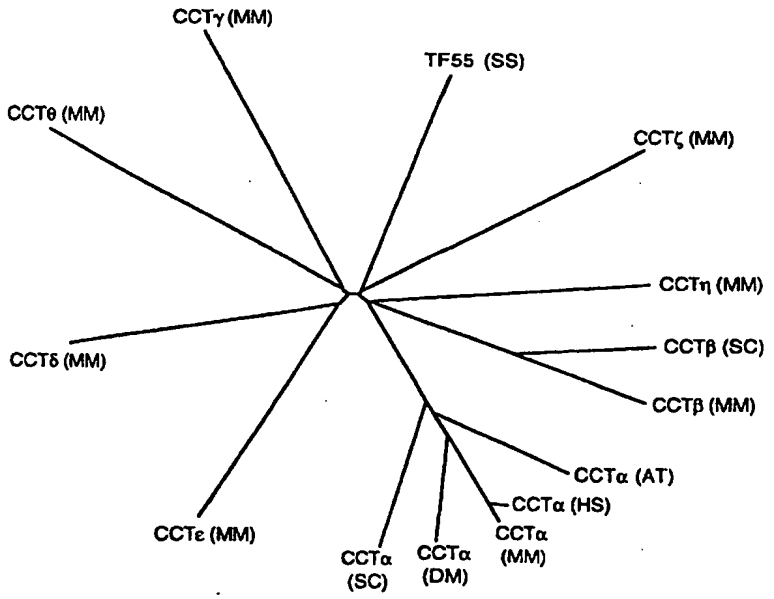
(74) Agents: WALTON, Seán, M. et al.; Mewburn Ellis, York House, 23 Kingsway, London WC2B 6HP (GB).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: FOLDING PROTEINS



(57) Abstract

Nucleic acid encoding subunits of a polypeptide folding complex containing TCP-1 is useful in expression of the subunits for assembly of the complex *in vitro* or *in vivo*. The complex promotes folding of polypeptides, such as tubulin, e.g. following recombinant expression. The sequences share a high degree of homology but the C-terminii are sufficiently different to enable peptides with sequences derived therefrom to be used in obtaining antibodies specific for the various subunits, i.e. able to distinguish between them. The antibodies may be used in analysis and design of complexes able to fold different polypeptides.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

FOLDING PROTEINS

The present invention relates to the folding of proteins. In particular, the present invention provides proteins, which are able to form a complex *in vitro*, useful in facilitating folding of proteins, for example those produced using recombinant DNA technology. Genes encoding the proteins of the complex are also provided. The present invention further relates to methods of assembling a protein complex able to fold proteins in an *in vitro* environment.

Molecular chaperones are known to be able to assist in the folding of proteins along the folding pathways from denatured state to correctly folded product (reviewed in [1]). One well studied class of molecular chaperones are the chaperonins (GroEL, Hsp60 and Rubisco subunit binding protein) found in eubacteria, mitochondria and plastids (reviewed in [2,3]). They are 14-mer double-torus structures composed of one (GroEL and Hsp60) or two (Rubisco subunit binding protein) subunit types and show seven-fold symmetry [4,5]. *In vitro* these chaperonins bind denatured proteins and upon ATP hydrolysis, release them into aqueous solution where they complete folding [6]. *In vivo* there is evidence that they are involved in the folding, transport and assembly of newly synthesized proteins. The original mutations isolated in groE affected the folding of bacteriophage particle subunits [7] but more recent genetic analysis suggests a more general role in protein biogenesis in *E.*

coli [8]. Hsp60 is involved in the import of proteins into the mitochondrial lumen from the cytoplasm [9].

Although no GroEL-like chaperonins have been identified in eukaryotic cytosol, the double-torus TCP-1-containing particle seems to be a component of the eukaryotic folding machinery and may play an analogous role to that of GroEL in eubacteria and the GroEL related chaperonins in symbiotic organelles. TCP-1 is weakly related to the GroEL family [10] but shows nearly 40% identity to an archaebacterial chaperonin, TF55 [11]. It has been proposed that GroEL and TCP-1 are subfamilies derived from a primordial gene [10,12,13,14] and it has been suggested that the eukaryotic TCP-1-containing chaperonin may have evolved from an archaebacterial lineage [3,11].

Recently, purified chaperonin containing TCP-1 has been shown to facilitate the folding of actin [15] and tubulin [16,17] *in vitro* and it binds newly-synthesized actin, tubulin and some other unidentified polypeptides *in vivo* [18]. One striking difference between the bacterially derived chaperonins and the TCP-1-containing chaperonin is the heteromeric nature of the TCP-1-containing particle [14,15,17,19]. There are at least five polypeptide species in the complex containing TCP-1 [14,17].

To date, little sequence information has been available on the polypeptides which make up the complex, despite the fact that various parties have obtained the

sequences of peptides from a number of polypeptides of TCP complexes of different organisms. Frydman et al (17) demonstrated the presence of six subunits in bovine TCP complex, which they termed "TRiC" (TCP-1 ring complex), and obtained some peptide sequence information indicating some resemblance both between polypeptides of the complex and between these polypeptides and those of other organisms. Rommelaere et al [59] looked at the cytosolic chaperonin from both rabbit reticulocyte lysate and bovine testis. They report finding eight different polypeptides in rabbit reticulocyte chaperonin, and obtaining partial amino acid sequences of all eight.

However, full length clones have proved elusive. The full sequence of murine TCP-1 has been available since 1986 (20) and Ehmann et al (FEBS, **336**: 2, 313-316, 1993) have reported the obtention of a TCP-1 related sequence from *Avena sativa* (oat) seedlings. Despite this information being available, there has yet to be a report of the obtention of full-length nucleic acid sequences encoding the components of a mammalian TCP-1 complex. Knowledge of short peptide sequences derived from individual subunits of chaperonin containing TCP-1 has not enabled the specific cloning of the full-length cDNA for individual subunits. One problem is that in order to be sure a peptide sequence is derived from a subunit of the family, it must be identifiably homologous to the only full-length mammalian sequence available, ie TCP-1 (20,48). Any DNA sequences derived by reverse

translation from the novel peptide sequence will also be related to TCP-1 and the related gene sequences. If these sequences are used as PCR primers they will prime synthesis and amplification of many TCP-1 related sequences, so further insight and activity are needed to identify the sequences which encode particular subunits of the complex.

The present invention provides individually seven nucleic acid molecules with sequences encoding subunits of the TCP-1-containing chaperonin, different from the original *Tcp-1* gene (reported in 20). Since, in mice, at least three of the novel *Tcp-1* related genes are unlinked to the mouse *t* complex, it is proposed to rename the TCP-1 complex [14] as CCT, chaperonin containing TCP-1. Only now we have all the eight complete sequences of the ubiquitously expressed subunits is it possible to know the gene and subunit to which each PCR product corresponds. Likewise, all other TCP-1 related genes in the databases make no sense without the complete sequences being available.

The present invention also provides molecules which are mutants, derivatives or alleles of any one of the seven sequences provided, particularly mutants, derivatives and alleles which encode a protein which retains a functional characteristic of the protein encoded by the respective wild-type gene, especially the ability to associate with at least another subunit to form a complex able to fold a polypeptide. Changes to a

sequence, to produce a mutant or derivative, may be by one or more of insertion, deletion or substitution of one or more nucleotides in the nucleic acid, leading to the insertion, deletion or substitution of one or more amino acids. Of course, changes to the nucleic acid which make no difference to the encoded amino acid sequence are included. We have demonstrated the existence of the 8 gene sequences in yeast and in plants by hybridization (52, unpublished results) and 6 genes have been isolated from yeast by ourselves and others (53). These 6 yeast genes correspond exactly (ie. they are the orthologues of) to six of the genes exemplified in this application CCTα, β, γ, δ, ϵ, η and ζ. We predict that all eukaryotic organisms contain at least the set of eight genes which we have described in mouse. There may be tissue specific CCT genes or additional CCT genes in some organisms but we would expect each of these to be closely related (greater than 70% amino acid sequence homology) to one of the eight genes described here. These eight CCT genes of mouse constitute the basic family which comprise the core CCT complex. In a preferred embodiment of the present invention, the sequence is one encoding a polypeptide found in a human or a mouse.

The polypeptides may have an amino acid sequence which shares a significant degree of homology with any of the specific sequences provided herein. Such homology may for example, be 60% or greater, 70% or greater, 80% or greater, 90% or greater or 95% or greater, provided

the polypeptide is able to function as a subunit of a complex able to fold a polypeptide.

The sequences of polypeptides encoded by nucleic acid according to each of seven different embodiments of the present invention are provided in Figure 3, (a) to (f) and Figure 8 (h) Preferred nucleic acid sequences are shown in Figure 8 (b) to (h). The present invention also provides a vector which comprises nucleic acid with any one of the provided sequences, preferably a vector from which polypeptide encoded by the nucleic acid sequence can be expressed. The invention further encompasses a host cell transfected with such a vector. The host cell may be bacterial or eukaryotic, such as yeast or mammalian eg murine.

According to another aspect of the present invention there is provided a method of making a polypeptide component (subunit) of a complex, which complex is able to fold a polypeptide, which method comprises expressing the polypeptide component from encoding nucleic acid, the component being other than TCP-1. (A nucleic acid sequence encoding TCP-1 is shown in Figure 8(a).) Preferably the nucleic acid is one with a sequence shown in Figure 8 or a mutant, derivative or allele of such a sequence. Appropriate cofactors or accessory materials may be needed in order for the complex to function in some circumstances. Indeed, specificity of the complex for folding different polypeptides may be altered according to which cofactors are present and/or what

combination of polypeptide components ("subunits") is used.

The present invention also provides a polypeptide component of a complex able to fold a polypeptide, the polypeptide being substantially free of other polypeptides or substantially free of other polypeptides of the complex, and being other than TCP-1. Preferably, the polypeptide component, or a mutant or derivative thereof can be found in a mammal, most preferably in a mouse or a human. In a preferred embodiment of the present invention the polypeptide is one which has any one of the amino acid sequences shown in figure 8 (b) to (h) (see also Figure 3). Modified polypeptides (eg mutants or derivatives) which have the ability to form a complex able to fold a polypeptide are encompassed by the present invention.

The provision for the very first time of full-length sequences of nucleic acid molecules which encode polypeptides of a complex able to fold a polypeptide enables the production of such a complex using recombinant techniques. According to another aspect of the present invention there is provided a method of producing a complex able (under appropriate conditions) to fold a polypeptide, the method comprising expressing polypeptide components of the complex from encoding nucleic acid and causing or allowing assembly of the polypeptide components into the complex. Preferably at least some of the components are murine or human.

Preferably a polypeptide component of the complex is any one of those shown in figure 8 (a) to (h), or a mutant or derivative thereof. Polypeptide components of the complex may be expressed individually and then purified, with the assembly carried out *in vitro* by mixing CCT subunits in the appropriate combinations. On the other hand, subunits may be expressed together.

Also provided by the present invention is a complex able to fold a polypeptide, the complex having been produced using recombinant DNA technology. The use of recombinant DNA technology in the production of a complex enables an intelligent selection to be made about which subunits to include, in those cases where not all of those provided are not required, and the easy assembly and purification of complexes with the required biological activity. This is facilitated still further by the provision of subunit specific antibodies, discussed below. The functional test for a useful complex is its ability to fold a polypeptide. Results shown below indicate that different combinations of subunits exist *in vivo*. Immunoprecipitation experiments may be used to determine subunit combinations in purified complexes. Furthermore, recombinant production of different combinations of subunits and *in vitro* testing of ability to bind substrate polypeptide (eg in a reticulolysate system) or ability to fold polypeptide (eg in an *E. coli* expression system or in a separate combination stage following earlier production of the

subunits) enables easy determination of subunit combinations which are functional.

The present invention also encompasses the use of a complex in the folding of a polypeptide. Generally the complex will be one produced using recombinant DNA technology. As discussed, the complex may consist of all or less than all of the subunits provided herein. Similarly, the invention provides a method of folding a polypeptide which comprises causing or allowing such a complex to fold the polypeptide, following a previous step of production of the complex. In fact, it may be that the complex is produced at the same time as the polypeptide to be folded, eg by expression from the same vector and/or expression in the same host cell, rather than in a previous step.

The provision by the present invention of the full-length sequences of the subunits of a folding complex enables the production of subunits individually, with subsequent purification and combining to form a complex, the production of subunits together with combination of subunits taking place in the host cell culture (eg *E. coli*, yeast or Baculovirus systems) and with subsequent purification, or the expression of combinations of subunits with substrate in a host cell system (eg *E. coli*, yeast or Baculovirus systems), thus constructing a folding factory *in vivo*.

Until now, it has proved problematic to obtain antibodies which are specific for an individual subunit

from the complex, in the sense of distinguishing between one subunit and the others. The reasons for this are apparent: the subunits share (varying degrees of) homology with one another. Accordingly, attempts to obtain subunit specific antibodies have failed. Immunisation of an animal with a purified subunit causes the production of antibodies which, in the most part, are cross-reactive with a number of subunits. Now that the present application has provided the full sequences of the subunits for the first time, it has proven possible to identify regions of the subunits which are sufficiently different from corresponding regions in the other subunits to enable the production of subunit specific antibodies.

Thus, according to another aspect of the present invention there is provided an antibody specific for a CCT subunit (subunit of a complex able to fold a protein) other than TCP-1 (CCTα), preferably a mammalian subunit, such as human or murine. In a preferred embodiment, the antibody is specific for an epitope of a peptide corresponding to a C-terminal portion of a subunit, or a variant of such a peptide (modified by any of insertion, substitution or deletion of one or more amino acids). Most preferred peptides for this purpose are as follows:

APRKRVPDHHPC	(CCTbeta)	SEQ. ID NO. 1
NRQTGAPDAGQE	(CCTgamma)	SEQ. ID NO. 2
SILKIDDVVNTR	(CCTdelta)	SEQ. ID NO. 3
IDDIRKPGESEE	(CCTepsilon)	SEQ. ID NO. 4

SAGRGRGQARFH	(CCTeta)	SEQ. ID NO. 5
SGKKDWDDDQND	(CCTtheta)	SEQ. ID NO. 6
EIMRAGMSSLKG	(CCTzeta)	SEQ. ID NO. 7

These peptides (including any variant modified in any way, provided the subunit specific nature of an antibody raised to the variant is retained) each represent an aspect of the present invention.

According to a further aspect of the invention there is provided a method of obtaining an antibody specific for a CCT subunit, other than TCP-1 (CCTα), which method comprises immunising a mammal (eg mouse, rat, rabbit, horse, goat, sheep or monkey) with a peptide of the subunit. Preferably, the peptide is one of those listed above, or a peptide which is a variant of one of those listed, provided subunit specificity of an antibody which binds the peptide variant is retained. Antibodies may be obtained from immunised animals using a variety of techniques known in the art, and screened, preferably using binding of antibody to antigen of interest (one of the C-terminal peptides, for example). For instance, Western blotting techniques or immunoprecipitation may be used (Armitage et al (1992) Nature **357**: 80-82).

As an alternative or supplement to immunising a mammal with a peptide, an antibody specific for a subunit may be obtained from a recombinantly produced library of expressed immunoglobulin variable domains, eg using lambda bacteriophage or filamentous bacteriophage which display functional immunoglobulin binding domains on

their surfaces; for instance see WO92/01047. The library may be "naïve", that is constructed from sequences obtained from an organism which has not been immunised with any of the peptides, or may be one constructed using sequences obtained from an organism which has been exposed to the peptide or peptides of interest.

Antibodies according to the present invention may be modified in a number of ways. Indeed the term "antibody" should be construed as covering any binding member having a binding domain with the required specificity. Thus the invention covers antibody fragments, derivatives, functional equivalents and homologues of antibodies.

Example antibody fragments, capable of binding an antigen or other binding partner are the Fab fragment consisting of the VL, VH, Cl and CH1 domains; the Fd fragment consisting of the VH and CH1 domains; the Fv fragment consisting of the VL and VH domains of a single arm of an antibody; the dAb fragment which consists of a VH domain; isolated CDR regions and F(ab')2 fragments, a bivalent fragment comprising two Fab fragments linked by a disulphide bridge at the hinge region. Single chain Fv fragments are also included.

A derivative is a substance derived from a polypeptide or antibody. The derivative may differ from a polypeptide from which it is derived by the addition, deletion, substitution or insertion of one or more amino acids, or the linkage or fusion of other molecules to the polypeptide. Changes such as addition, deletion,

substitution or insertion may be made at the nucleotide or protein level.

A hybridoma producing a monoclonal antibody according to the present invention may be subject to genetic mutation or other changes. It will further be understood by those skilled in the art that a monoclonal antibody can be subjected to the techniques of recombinant DNA technology to produce other antibodies or chimeric molecules which retain the specificity of the original antibody. Such techniques may involve introducing DNA encoding the immunoglobulin variable region, or the complementarity determining regions (CDRs), of an antibody to the constant regions, or constant regions plus framework regions, of a different immunoglobulin. See, for instance, EP184187A, GB 2188638A or EP-A-0239400. Cloning and expression of chimeric antibodies are described in EP-A-0120694 and EP-A-0125023.

The peptides for use in obtaining antibodies may be made by any of a variety of techniques known to those skilled in the art. For instance, this may be by solid phase synthesis (Merrifield, JACS 85: 2159-2154 (1963)) or by techniques described in Bodanszky et al, *Peptide Synthesis*, second edition (Wiley, 1976). Standard solution peptide synthesis methodologies, using chemical or enzymatic methods of amide bond formation may be employed. Commercial peptide synthesizing machines are available.

Conveniently, an amino terminal cysteine can be added to the chosen peptide in order that the peptide can be coupled to PPD through the cysteine using standard chemistry.

Systems for cloning and expression of a polypeptide in a variety of different host cells are well known. Suitable host cells include bacteria, mammalian cells, yeast and baculovirus systems. Mammalian cell lines available in the art for expression of a heterologous polypeptide include Chinese hamster ovary cells, HeLa cells, baby hamster kidney cells and many others. A common, preferred bacterial host is *E. coli*.

Suitable vectors can be chosen or constructed, containing appropriate regulatory sequences, including promoter sequences, terminator fragments, polyadenylation sequences, enhancer sequences, marker genes and other sequences as appropriate. For further details see, for example, *Molecular Cloning: a Laboratory Manual*: 2nd edition, Sambrook et al, 1989, Cold Spring Harbor Laboratory Press. Transformation procedures depend on the host used, but are well known.

Baculovirus expression systems are commercially available in kit form, eg MaxBac™. The techniques employed are described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin* 1555 (1987). See also *The Molecular Biology of Baculoviruses* (ed Doerfler, 1986) for a general overview of Baculovirus. Baculovirus expression vectors have been developed for

infection into a number of cell-types, including those from *Aedes aegypti*, *Autograph californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda* and *trichoplusia ni*.

Embodiments of the present invention are now described further by way of illustration, with reference to the figures identified below. All documents mentioned in the text are hereby incorporated by reference.

Figure 1 shows one- and Two-dimensional gel electrophoresis of mouse CCT and bovine TRiC. (a) SDS-PAGE analysis (8% gel) of mouse testis CCT. TCP-1-containing particles purified by sucrose gradient fractionation followed by ATP-affinity column chromatography from mouse testis [14] were electrophoresed. (b-c) 2D-PAGE analysis of mouse CCT (same sample as Figure 1a) and bovine TRiC [17] (kind gift of F.-U. Hartl). Subunits of TCP-1-containing particles purified by sucrose gradient fractionation followed by ATP-affinity column chromatography from mouse testis (b) and bovine testis (c) were separated by IEF followed by SDS-PAGE. (d) 2D-NEPHGE analysis of bovine TRiC. Subunits of bovine TRiC were separated by NEPHGE followed by SDS-PAGE. (e) 2D-NEPHGE analysis of antibody-affinity purified CCT. TCP-1-containing particles were purified from ^{35}S-labelled F9 cell extracts using monospecific monoclonal antibodies to TCP-1. Proteins were visualized by silver staining (a-c), Coomassie staining (d) or autoradiography (e). Arrowhead

indicates a 45 kDa protein of pI 6.5 which co-purified with F9 cell CCT. Arrows show Hsp70 proteins.
(f) Molecular weights and pI values of the polypeptides comprising mouse CCT and bovine TRiC. These values are determined from 2D-gel analysis with both pI and molecular weight markers.

Figure 2 shows an alignment of peptide sequences deduced from 5'-end DNA sequences of 16 *Tcp-1* related genes (SEQ. ID NO's 8-23). Clone names are used except for *Tcp-1* genes of mouse, human and yeast in this figure. Letters in parentheses indicate the source species: M, *Mus musculus* (mouse); H, *Homo sapiens* (human); C. *Caenorhabditis elegans* (nematode); S, *Saccharomyces cerevisiae* (yeast). Sources of the nucleotide sequences are as follows: mouse *Tcp-1* [20,48]; human *Tcp-1* [48]; yeast *Tcp-1* [27]; 4950 [28]; pG1-pG4 (partial 5'-end sequences) [22] and this report; p383 and p384 (partial 5'-end sequences), pTß2, pTγ7, pTδ2, pTε5, pTζ12 and pCBL80, this report. Although three additional human *Tcp-1* related partial DNA sequences, pAP3 (A. Malik *et al.*, manuscript in preparation), HTR3 [41] and IB713 [49] are known, these sequences are not added to this figure due to absence of 5'-end sequences (pAP3 and HTR3) or inaccuracy of sequence (IB713). However, judging by DNA sequence homology, IB713, pAP3 and pHTR3 are members of groups 2, 5 and 6, respectively.

Figure 3 shows amino-acid sequences of ß, γ, δ, ε, ζ and η subunits of mouse CCT. Amino-acid sequences of

mouse CCTß (a) (SEQ. ID NO. 25), CCTγ (b) (SEQ. ID NO. 26), CCTδ (c) (SEQ. ID NO. 27), CCTε (d) (SEQ. ID NO. 28), CCTζ (e) (SEQ. ID NO. 29) and CCTη (f) (SEQ. ID NO. 30) deduced from nucleotide sequences of cDNA clones pTß2, pTγ7, pTδ2, pTε5, pTζ12 and pCBL80 respectively. Sequences of tryptic peptides derived from B1-B4 of mouse CCT (Figure 1a) are shown by bold underlining. Thin underlines indicate amino acids identical to those of bovine P1-P3 and P5 ([17] and Figure 1) determined from tryptic polypeptides of the TRiC subunits ([17], and unpublished results of F.U. Hartl *et al.*). An amino-acid position not determined by peptide sequencing is indicated by an 'x'. A bovine tryptic peptide sequence, P5/T36, contains 3 amino-acid differences compared to the mouse sequence and these are shown underneath.

Figure 4 shows alignment of amino acid sequences of eight mouse CCT subunits and the homo-oligomeric archaebacterial chaperone TF55. A dash indicates an amino-acid gap. Conserved amino acids are indicated by bold characters. Consensus amino acids of these proteins are shown under the alignment (SEQ. ID NO. 33). The amino acid sequence of CCTθ is SEQ. ID NO. 31). The amino-acid sequences of mouse CCTα (TCP-1) [20,48] (SEQ. ID NO. 24) and archaebacterial chaperone TF55 of *Sulfolobus shibatae* [11] (SEQ. ID NO. 32) are derived from previous publications.

Figure 5 shows comparison between a consensus motif of CCT subunits and the ATP binding motif of cAMP

dependent kinase and other kinases in this family. Amino-acid sequences around a consensus motif of CCT subunits (positions 102-115 in Figure 4) and the ATP binding motif of cAMP dependent kinase and related kinases [25,26] are compared. Amino acids conserved between these two groups are indicated by bold characters. (CCTα-SEQ. ID NO. 34, CCTβ-SEQ. ID NO. 35, CCTγ-SEQ. ID NO. 36, CCTδ-SEQ. ID NO. 37, CCTε-SEQ. ID NO. 38, CCTζ-SEQ. ID NO. 39, CCTη-SEQ. ID NO. 40, CCTθ-SEQ. ID NO. 41) cAPK-α, cyclic-AMP-dependent kinase α, (SEQ. ID NO. 42); PKC-α, protein kinase c-α (SEQ. ID NO. 43); CaMII-α, bovine calcium-calmodulin-dependent kinase type IIα, (SEQ. ID NO. 44) SNF1, a budding-yeast wild-type gene product for sucrose non-fermenting mutant (SEQ. ID NO. 45); cdc2+, a fission-yeast cell-division-cycle gene product (SEQ. ID NO. 46); CDC7, a budding-yeast cell-division-cycle gene product (SEQ. ID NO. 47); Raf, a human cellular homologue of murine sarcoma virus product (SEQ. ID NO. 48); Src, a human cellular homologue of Rous avian sarcoma virus product (SEQ. ID NO. 49); Abl, a human cellular homologue of murine leukaemia virus product (SEQ. ID NO. 50); EGFR, human epidermal growth factor receptor (SEQ. ID NO. 51); INSR, human insulin receptor (SEQ. ID NO. 52); PDGFR, mouse platelet-derived growth factor receptor (SEQ. ID NO. 53).

Figure 6 shows hybridization analysis of the *Cct* genes of mouse and yeast. (a) Southern analysis of mouse *Cct* genes. Mouse 129/Sv liver DNA (10 μg/lane) was

digested with HindIII, electrophoresed on a 0.7% agarose gel and blotted onto a nylon membrane. The membrane was cut into 7 pieces and each of them hybridized with ^{32}P-labelled mouse *Ccta/Tcp-1* (clone pT1b11) [33], *Cctb* (pTß2), *Cctg* (pTγ7), *Cctd* (pTδ2), *Ccte* (pTε5), *Cctz* (pTζ12) and *Ccth* (pCBL80) 1.5kb cDNA probes. These membranes were washed in 0.1x SSC supplemented with 0.1% SDS at 65°C. (b) Southern analysis for yeast homologues of *Cct* genes with mouse cDNA probes. *Saccharomyces cerevisiae* DNA (1.5 μg/lane) was digested with PstI. The DNA was electrophoresed, blotted and hybridized as described in panel (a). After the hybridization, these membranes were washed in 2 x SSC supplemented with 0.1% SDS at 58.5°C. Positions of molecular weight markers are shown on the left side of each panel.

Figure 7 shows the evolutionary tree of CCT subunits of eukaryotes and the homo-oligomeric chaperonin of the archaebacterium *Sulfolobus shibatae*, TF55. An evolutionary tree based on amino acid substitutions is constructed by the neighbour-joining method [50] with the aligned amino-acid sequences numbers 18-67, 70-157, 169-200, 213-241, 251-272, 274-310, 326-372, 376-388, 398-505 and 521-566 (472 amino acids for each subunit) of mouse CCT subunits and archaebacterial chaperone TF55 in Figure 4, and corresponding amino acid sequences of other homologues. Amino-acid sequences of human [48], fruit fly [51], plant [29] and yeast [27] homologues of CCTα/TCP-1 were derived from previous publications. The

amino-acid sequence of the yeast homologue of CCTß was obtained from Miklos *et al.* [28]. Capitals in parentheses indicate source species: MM, *Mus musculus* (mouse); HS, *Homo sapiens* (human); DM, *Drosophila melanogaster* (fruit fly); CE, *Caenorhabditis elegans* (nematode); AT, *Arabidopsis thaliana* (plant); SC, *Saccharomyces cerevisiae* (yeast); SS, *Sulfobus shibatae* (archaebacterium).

Figure 8 shows full nucleotide sequences encoding subunits of a complex, in this case mouse *CCT* cDNAs. The polypeptide sequences of the predicted open reading frames are given under each DNA sequence. Figure 8(a) is the sequence of Tcp-1a/*Ccta* (SEQ. ID NO. 54). Figures 8 (b) (SEQ. ID NO. 55), (c) (SEQ. ID NO. 56), (d) (SEQ. ID NO. 57), (e) (SEQ. ID NO. 58), (f) (SEQ. ID NO. 59), (g) (SEQ. ID NO. 60) and (h) (SEQ. ID NO. 61) are the sequences of the *Cctb, Cctg, Cctd, Ccte, Cctz, Ccth and Cctq* genes respectively.

Figure 9 shows immunoprecipitation of CCT complexes from mouse F9 cells, precipitated under native conditions (14) using anti-peptide antibodies specific to CCTß, CCTγ, CCTε and CCTη. A negative control shows absence of signal using an anti-peptide antibody to *Schizosaccharomyces pombe* CCTα, which does not recognise any mammalian subunits.

Figure 10 2D-PAGE analysis of CCT. Subunits of ATP-affinity purified CCT from mouse testis were separated by 2D-PAGE and proteins were visualised by silver staining

(A). Greek letters indicate the eight subunit species whose genes have been cloned, and S6 shows the testis-expressed subunit. A novel co-purifying 63kDa protein of pI 6.93 is indicated by an arrow and hsp 70 proteins are shown by arrowheads. CCT subunits were immunoblotted with rabbit antibodies against carboxy terminal amino acid sequences of CCTθ (B), CCTϵ (C), CCTβ (E), CCTγ (F), CCTζ (G), CCTη (H), CCTδ (I), and a monoclonal antibody, 91a, against CCTα (D). Panel J shows CCT subunits immunoblotted with a rabbit antibody to a chaperonin consensus sequence thought to be involved in ATP-binding. Sequences of the peptide immunogens are shown in Table 2. In all panels the acidic side is to the left.

Figure 11 Analysis of native populations of CCT. Partially purified CCT from mouse testis was subjected to non-denaturing isoelectric focusing followed by SDS-PAGE. Proteins were visualised by silver staining (A). Native populations of CCT resolved by non-denaturing IEF are indicated by I and II and arrowheads show hsp 70 proteins. Native populations of CCT were immunoblotted with a rabbit antibody against CCTϵ (B), a 30kDa proteolytic fragment of CCTϵ is indicated by an arrow, or a monoclonal antibody against β-tubulin (C). In all panels the acidic side is to the left.

FIGURE 12 shows *in vitro* translation products produced from individual CCT cDNA clones. Analysis of 35 S-labelled CCT subunits synthesized in rabbit reticulocyte lysate from individual CCT cDNA clones.

Aliquots of the *in vitro* translation reactions are electrophoresed on 10% SDS-PAGE and visualized by autoradiography. Lanes: (1) CCTα, (2) CCTβ, (3) CCTγ, (4) CCTδ, (5) CCTϵ, (6) CCTζ, (7) CCTη, (8) CCTθ. M- molecular weight markers.

FIGURE 13 shows detection of labelled CCT complex.

Figure 13A shows an autoradiogram of a 10% SDS-polyacrylamide gel of 35S-labelled CCT proteins detected in fractions from a sucrose density grandient analysis of CCT assemblec *in vitro* in rabbit reticulocyte lysate. Eight CCT subunit mRNAs, CCTα-CCTθ were translated together and the reaction mix applied to a 10%-40% sucrose gradient as previously described (14). Lane M = marker proteins, lanes 1-14 contain 30 microlitre aliquots of 1ml sucrose fractions (fractions 17-4). Fraction 17 is the first, lightest fraction of the gradient. Free CCT subunits are observed in fractions 16 and 15 (lanes 2 and 3) and the assembled complex is found in fractions 12 and 11 (lanes 6 and 7).

Figure 13B shows an autoradiogram of a 10% SDS-polyacrylamide gel of 35S-labelled CCT proteins immunoprecipitated from sucrose gradient fractions by anti-CCTβ antibodies. This analysis demonstrates the presence of assembled CCT in sucrose fractions 12 and 11, but not in fractions 16 and 15. Immunoprecipitation of aliquots of fractions from the same sucrose gradient as in Figure 13A. 300 microlitre aliquots of each of gradient fractions 17-9 were immunoprecipitated in a 1ml

reaction containing 5 microlitres of anti-CCTβ antibodies. Immune-complexes were recovered by binding to protein-A sepharose beads. Precipitation of fractions 16 and 15 with anti-CCTβ antibodies only recovers CCTβ from the mixture of 8 free, unassembled CCT subunits in the light fractions. However, precipitation of fractions 12 and 11 recovers all subunits consistent with the notion that these sucrose fractions contain assembled CCT of expected size and sedimentation properties.

Summary of materials and methods used in obtaining gene sequences.

Purification of CCT

Mouse testis CCT was purified by sucrose gradient fractionation followed by ATP-agarose affinity chromatography as previously described [14]. Bovine testis TCP-1 ring complex (TRiC) [17] was a kind gift of F-U Hartl. Mouse F9 cells were labelled with ^{35}S-methionine and ^{35}S-cysteine for 5h and CCT was purified by immunoaffinity chromatography using anti-TCP-1 monoclonal antibodies 84A and 91A [14,21]. Isoelectric focusing (IEF) [43] and non-equilibrium pH gradient electrophoresis (NEPHGE) [44] were carried out according to Corbett and Dunn [45].

Peptide sequencing of CCT subunits

CCT was subjected to electrophoresis on an 8% SDS-PAGE gel and protein bands were stained with 0.1% Coomassie Brilliant Blue in 0.5% acetic acid/10% aqueous

methanol and destained with 10% aqueous methanol. Bands as shown in Figure 1a were excised and the proteins were electroeluted from the gel slices. The proteins were concentrated onto polyvinylidene difluoride (PVDF) membranes using ProSpin cartridges (ABI) and each sample was digested with trypsin for 18h at 37°C using the method described by Fernandez *et al.* [46]. Digestion was terminated with trifluoroacetic acid (TFA) and tryptic peptides were isolated from the PVDF supernatants using an ABI 130A microbore separation system. TFA buffers (A: 0.1% TFA; B: 0.085% TFA, 70% acetonitrile) and a simple linear gradient (5 to 100% B in 50 min; 200 µl per min) were employed on a 2.1 x 100 mm Brownlee Aquapore RP-300 column. Sequence analysis on the isolated fragments was performed using an ABI 477A protein sequencer with on-line PTH detection and data handling.

Eight TRiC peptide sequences were published by Frydman *et al.* [17] and a further TRiC peptide sequence (P2) was a kind gift of F-U. Hartl.

C. elegans Tcp-1 related cDNA clones

Nine *Tcp-1* related cDNA clones were reported from 5'-end single-pass sequencing of randomly selected *C. elegans* cDNA clones [22]. We compared these 9 partial sequences and found that they were derived from 4 independent genes. Because the cDNAs were all directionally cloned, we subcloned the cDNA having the longest 5' region for each of the four genes (cm08g10, cm12b10, cm11d3 and cm13e8). These pBluescript KS$^+$ clones

were renamed pG1, pG2, pG3 and pG4 and represent 1.8 kb inserts for each of these four *C. elegans* genes. We accurately sequenced the 5' ends of these plasmid subclones and deduced 90 amino acids of NH_2-terminal sequence for each gene (Figure 2). The deduced peptide sequences of pG1, pG2, pG3 and pG4 have 74%, 31%, 52% and 48% identity to mouse TCP-1 [20] at their NH_2 termini.

Human Tcp-1 related cDNAs

A cDNA library of the human cell line HT1080 (kindly given by P. Mitchell) contains cDNA inserted at the EcoRI site of bacteriophage lambda ZAPII. These phages (1200 pfus) were plated onto each of 24 plates and phage pools of each of them were made by incubating each lawn in 5 mls of SM buffer overnight. Approximately 1.5 mls of pooled phage stock was obtained from each plate. A mixed primer for the peptide sequence TNDGATI (a motif conserved between TCP-1 and TF55) was designed (SEQ. ID NO. 62):

5'-AT(AG)GT(AGT)GC(ACT)CCATC(AG)TT(AGT)GT-3'.

In and out PCR using the specific primer and a primer of the lambda ZAPII polylinker on the 5' side of the inserts was performed. PCR reactions (60 mM KCl, 15 mM Tris-HCl pH8.8, 2.25 mM $MgCl_2$, 250 μM each dATP, dGTP, dCTP and dTTP, 0.4 pmol/μl each primer, 25 μl/tube) were prepared and then 1 μl of the phage stock and 1.5 unit Taq polymerase was added to each tube. These were subjected to 30 cycles of 93°C for 15 sec, 45°C for 15 sec and 72°C for 30 sec. PCR products from 20 of the

pools gave one or only a few 150-300 bp bands by 1% agarose gel electrophoresis. These bands were excised from the gel, reamplified and sequenced. Three *Tcp-1* related sequences were identified, subcloned into pBluescript KS$^+$ and named p383, p384 and ph13. The deduced polypeptide sequence from p383 has 28% identity to mouse TCP-1 over its NH_2-terminal 54 amino acids (excluding the primed sequence TNDGATI (SEQ. ID NO. 63)), that from p384 has 41% identity to mouse TCP-1 over its NH_2-terminal 51 amino acids (excluding TNDGATI) and that from ph13 has 24% identity to mouse TCP-1 over its NH_2-terminal 65 amino acids (excluding TNDGATI).

A partial cDNA clone, pAP3, was isolated from an adult human kidney cDNA library by hybridisation with a PCR fragment generated using redundant primers corresponding to Na/K transporter ATPase sequences (Malik, manuscript in preparation). Upon sequencing, pAP3 turned out to contain a 1050 base pair insert encoding a *Tcp-1* related gene. Total mouse brain cDNA was subjected to PCR using two primers from the human pAP3 sequence:

forward 5'-ACATCCAGCAGCTCTGTGAG-3' (SEQ. ID NO. 64) and

reverse 5'-CCTTGCCTAGCACTCACTCC-3' (SEQ. ID NO. 65). An 800 bp fragment of the mouse orthologue of pAP3 gene was generated.

Isolation of mouse Tcp-1 related cDNA clones

The mouse full-length cDNA clones encoding 6 novel

TCP-1 related proteins were cloned as follows. An F9 embryonal carcinoma cell cDNA library was made in lambda ZAP by the method of Nagata *et al.* [47] and $7.5x10^4$ size-selected recombinants were obtained from 1 μg poly(A)$^+$ RNA. This library was transferred onto GeneScreen Plus (NEN) membranes and duplicate filters were sequentially probed with the following five *Tcp-1* related cDNA probes. The two *C. elegans* cDNAs, pG2 and pG3, the three human cDNAs, p383, p384 and ph13 and the 800 bp mouse RT-PCR product from pAP3 were ^{32}P-labelled and hybridized to the membranes in Southern hybridization (SH) buffer (6x SSPE, 5x Denhart's reagent, 0.5% SDS, 10 mg/ml salmon sperm DNA) overnight at 55°C for the two *C. elegans* probes, 60°C for the three human probes, and 65°C for the mouse probe. Between 6-12 positively hybridizing clones were purified for each probe. The longest clone (1.8-2.2 kb inserts) derived from each screening was named as pTß2 (p383 homologue), pTε5 (p384 homologue), pTθ1 (ph13 homologue), pTζ12 (pG2 homologue), pTδ2 (pG3 homologue) and pTγ7 (pAP3 homologue). The relationship between these probes and clones is described in Figure 2. The sixth mouse *Tcp-1* related cDNA was recovered during a pilot experiment for a mouse testis cDNA sequencing project. Ninety-five random mouse testis cDNAs directionally cloned in lambda ZAPII were single pass sequenced from their 5′ ends. The sequence of the 80th cDNA revealed a new *Tcp-1* related gene which is different to all the others described above and this plasmid,

pCBL80, was completely sequenced.

DNA sequencing and computer analysis

The nucleotide sequences of all the *Tcp-1* related genes, except pCBL80, were determined by the dideoxynucleotide chain termination method with fluorescently labelled primers using the PRISM kit and a 373A automated sequencer (ABI). Clone pCBL80 was manually sequenced by dideoxynucleotide chain termination and ^{35}S-dATP using a combination of deletion clones and specific primers. All these sequences and deduced amino-acid sequences were analysed by UWGCG programs on the Silicon Graphics Crimson network at the SERC, Daresbury, UK. Phylogenetic analyses were also carried out at Daresbury using the PHYLIP collection of programs.

Southern blotting

Mouse strain 129/Sv DNA (10 μg/lane) and *S. cerevisiae* strain 3a DNAs (1.5 μg/lane) were digested with HindIII and electrophoresed on 0.7% agarose gels and blotted onto GeneScreen Plus membranes. Probes of 1500 ± 3 bp length were made by PCR with the mouse *Tcp-1* cDNA (pT1b11) [33] and mouse *Tcp-1* related cDNAs (pTß2, pT 7, pTδ2, pTε5, pT 12 and pCBL80). These 7 probes were labelled with ^{32}P and hybridized to the blotted membranes in SH buffer overnight at 65°C (mouse) or at 55°C (yeast). The membranes were washed either in 0.1x SSC containing 0.1% SDS at 65°C (mouse) or in 2x SSC containing 0.1% SDS at 58.5°C (yeast).

Results

Subunit composition of the chaperonin containing TCP-1 (CCT)

Two-dimensional gel analysis of the subunits of the CCT reveals 7-9 spots of 52-65 kDa. Figure 1 shows SDS-PAGE, 2D-PAGE and 2D-NEPHGE comparison of CCT isolated from mouse testis and F9 embryonal carcinoma cells, and bovine testis TCP-1 ring complex (TRiC) [17].

Excluding the Hsp70 species which co-purify specifically with the CCT [14], the testis preparations contain 9 major species (Figure 1b-d) and the F9 cell preparation contains seven 53-65 kDa species excluding a 45 kDa protein (Figure 1e). This represents an increase in the number of species reported by Lewis *et al.* [14] because two species (S8 and S9), with basic pIs, resolve poorly in the IEF dimension of 2D-PAGE run to equilibrium. A 2D-NEPHGE analysis of bovine TRiC clearly shows 9 subunit species (Figure 1d).

We had previously shown, using a monoclonal antibody (91A) [21] which recognizes only a single subunit (S3, TCP-1) of the CCT, that antibody-affinity purified CCT has a similar electrophoretic profile to CCT purified by ATP-agarose affinity or ion-exchange chromatography, on SDS-PAGE gels [14]. Figure 1e shows a 2D-NEPHGE analysis of antibody-affinity purified, CCT from F9 cells; the polypeptide spot profile strongly resembles CCT biochemically purified from testis (Figure 1a,b) and seven 52-65 kDa species are distinguishable by the 2D analysis. A 45 kDa protein of pI 6.5 was co-purified

with CCT by the antibody-affinity method, suggesting a novel co-factor or substrate of CCT (arrow head in Figure 1e). These results confirmed that TCP-1 comprises a hetero-oligomeric complex with other polypeptides.

Figure 1f shows a tabulation of the apparent isoelectric points and molecular weights of the CCT subunits. Frydman *et al.* [17] reported that bovine testis TRiC migrated as 5 bands on a 12.5% SDS-PAGE gel (P1-P5). We found that band P4 contained 5 species of polypeptides which were distinguishable from one another on a 2D gel, and we named them P4.1, 4.2, 4.3, 4.4 and 4.5 according to ascending pI. The subunits of mouse CCT (S1-S9) are numbered according to ascending pI using the mouse testis pattern as the standard. These bovine and mouse spots are distributed from pH6.1 - pH7.1 and MW 65 kDa - MW52 kDa except for the most basic spot (NEPHGE analysis suggests a value of around pH 7.4 for the pI of this most basic spot).

Isolation of mouse cDNAs encoding seven novel TCP-1 related proteins

We previously reported [14] the existence, in various eukaryotes, of numerous genes and partial cDNA sequences related to the original mouse *Tcp-1* gene [20], the first identified gene encoding an CCT subunit [14,17]. It proved necessary to use a combination of methods to isolate murine *Tcp-1* related genes and clone the full-length cDNAs which encode the seven novel TCP-1 related proteins described herein. The first and second

(clones pTδ2 and pTζ12) were isolated by cross-hybridization with *Caenorhabditis elegans* cDNA probes recovered from the 5' expressed sequence tag collection of Waterston *et al.* [22].

The third, fourth and seventh (clones pTß2, pTε5 and pTθ1) were isolated by cross-hybridization with human probes recovered by PCR of human HT1080 cell line cDNA using degenerate primers from a conserved region of TCP-1 and TF55.

The fifth gene (clone pCBL80) was isolated during a mouse testis cDNA sequencing project at the Chester Beatty Laboratories, London. A cDNA fragment of a *Tcp-1* related gene (clone pAP3) was recovered accidentally from a human kidney cDNA library during a screen for ion transport channel genes (Malik *et al.*, manuscript in preparation) and the sixth murine gene (clone pTγ7) was recovered by hybridization with a mouse cDNA PCR product which was made with primers derived from the sequence of this human cDNA.

Figure 2 shows the very conserved NH_2-terminal domain deduced from these various *Tcp-1* and *Tcp-1* related genes of mouse, human, *C. elegans* and *Saccharomyces cerevisiae*. A phylogenetic analysis (UPGMA) [23] based on the amino-acid sequences of the domain showed seven groups of genes (data not shown), suggesting that in eukaryotes *Tcp-1* and the *Tcp-1* related genes have each evolved independently since they diverged. Analysis of the eight members of the *Cct* gene family suggests that the heteromeric CCT

particle had already evolved in a common ancestral organism of animals, plants and yeast more than 500 million years ago and that different subunits may have different functions in protein folding.

Figure 3 and Figure 8 show amino-acid sequences of the seven mouse TCP-1 related proteins deduced from their full-length cDNA clones.

Figure 8 (b) to (h) shows the DNA sequences of the seven genes and the amino acid sequences (of the deduced open reading frames) underneath.

The seven Tcp-1 related cDNAs encode CCT subunits

Internal tryptic peptides derived from an SDS-PAGE gel separation of mouse testis CCT (B1-B4, Figure 1a) were sequenced because the NH_2 termini of every subunit appeared to be blocked yielding no sequence data. Frydman *et al.* [17] obtained some internal peptide sequences from four of the bovine testis TRiC subunits (P1, P3, P4a/TCP-1 and P5) and they also made available to us unpublished peptide sequences which we show here correspond to TRiC subunit P2. Comparisons between these mouse and bovine peptide sequences and the predicted peptide sequences of the *seven Tcp-1* related cDNAs reveals that they each encode particular CCT subunit proteins (Figure 3, Figure 8 (h), Table 1). We named these 7 new genes as *Cctb* (pTß2), *Cctg* (pTγ7), *Cctd* (pTδ2), *Ccte* (pTε5), *Cctz* (pTζ12), *Ccth* (pCBL80) and *Cctq* (pTθ1) and their encoded proteins as CCT ß, γ, δ, ζ, η, and θ subunits respectively. We suggest renaming the

Tcp-1 gene [20] as *Cctα* and the TCP-1 protein as CCTα. Table 1 summarizes the correspondence between the proteins encoded by *Cct* genes and the spots of CCT subunits separated by 2D-gel electrophoresis (Figure 1).

Similarities among CCT subunit polypeptides

Table 1 also shows the lengths in amino-acid number and molecular weights of the CCT subunits. They have similar characteristics despite the fact that they are only around 30% identical to one another in amino-acid sequence (Table 2). They vary slightly in length from 531-556 residues and in predicted molecular weight from 57456-60636 Da. These predicted molecular weights are approximately consistent with those determined by SDS-PAGE (Figure 1).

The pIs of the subunits, experimentally determined by isoelectric focusing, correlate well with the total charge value of each predicted CCT subunit polypeptide assuming some histidine residues are charged as cations at their respective pIs (data not shown). The percentage of hydrophobic and charged amino acids in each CCT subunit is highly conserved, ranging from 31.6-33.5% and 23.9-27.3% respectively.

These conserved chemical properties probably reflect common functions of each mammalian subunit and are shared with the archaebacterial chaperonin TF55. Agard [24] suggested that hydrophobic interaction between the *E. coli* chaperonin GroEL and its substrates could be important for the folding process. The conserved

percentage of hydrophobic amino acids in CCT subunits may thus be important for interaction with substrates. Hydrophobicity profiles of CCT subunits suggest conservation of hydrophobic and hydrophilic amino-acid distributions among CCT subunits, especially at the NH_2 terminus (data not shown).

Figure 4 shows an alignment of all eight CCT sequences and TF55. The alignment of the peptide sequences show six major gaps (Nos. 164, 201, 246, 315, 392 and 508 in Figure 4) and thus indicates seven blocks of homology divided by the gaps. Amongst these seven blocks, the first (Nos. 1-164) and fifth (Nos. 392-508) are the largest and contain some highly conserved motifs: LGPKGMDKM (Nos. 52-60) (SEQ. ID NO. 66), TITNDGATIL (Nos. 71-80) (SEQ. ID NO. 67), QDDEVGDGTTSVV (Nos.100-112) (SEQ. ID NO. 68), ERSLHDAL (Nos.423-430) (SEQ. ID NO. 69), and VV(A/P)GGGA (Nos. 442-448) (SEQ. ID NO. 70). The third motif includes a motif absolutely conserved in CCT subunits, GDGTT (SEQ. ID NO. 71), previously recognized by Lewis *et al.* [14] to be homologous to a nucleotide phosphate binding domain of cyclic-AMP-dependent kinase and other members of this kinase family [25,26]. Figure 5 shows a comparison between these motifs. This suggests that all the CCT subunits share the common function of ATPase activity.

Structural genes encoding CCT subunits and homologous genes in yeast

Southern analysis of mouse genomic DNA (Figure 6a)

shows the independent structural genes encoding seven mouse CCT subunits. Hybridization of *Saccharomyces cerevisiae* DNA with seven mouse *Cct* cDNA probes (Figure 6b) shows one or two bands of PstI-digested DNA for each probe. Two *Tcp-1* related genes have already been identified in yeast; the orthologue of *Ccta/Tcp-1* [27] and the orthologue of *Cctb* [28]. Both these genes are essential [27,28] and temperature sensitive mutations in both genes affect microtubule mediated processes [28]. The CCT of *S. cerevisiae* is composed of numerous subunits and the yeast homologues of CCTα/TCP-1 and CCTß are components of the same complex since they co-purify together [28]. Preliminary sequence data derived from yeast DNA clones corresponding to the novel sequences detected with the mouse *Cct* probes suggests that they are indeed the orthologues of the mouse *Cct* genes. Southern analysis of *Arabidopsis* genomic DNA with the seven *Cct* cDNA probes also suggests the existence of plant homologues for each subunit (data not shown); the *Arabidopsis* homologue of *Ccta/Tcp-1* has been reported [29]. It is likely that these seven species of CCT subunits are ubiquitous in all eukaryotes.

Determination of the subunit compositions and arrangements in CCT

The sequence of the eight CCT subunits reveals that they are rather divergent at the C-terminii. A monoclonal antibody, 23C, had previously been made (21) which reacts with the C-terminus of CCTalpha (Harrison-

Lavoie et al, 1993, EMBO J. 7: 2847-2853) and can be used to precipitate CCT complexes. Subunit specific antibodies have now been made, utilising the sequence information provided by the invention. The antibodies were raised against C-terminus peptides with the following sequences:

CCTbeta:	APRKRVPDHHPC	(SEQ. ID NO. 1)
CCTgamma:	NRQTGAPDAGQE	(SEQ. ID NO. 2)
CCTdelta:	SILKIDDVVNTR	(SEQ. ID NO. 3)
CCTepsilon:	IDDIRKPGESEE	(SEQ. ID NO. 4)
CCTeta:	SAGRGRGQARFH	(SEQ. ID NO. 5)
CCTtheta:	SGKKDWDDDQND	(SEQ. ID NO. 6)
CCTzeta:	EIMRAGMSSLKG	(SEQ. ID NO. 7)

The peptides were synthesised by conventional solid phase chemistry on a commercial peptide synthesizing machine. An amino terminal cysteine was added to the peptides and used to couple them to purified protein derivative (PPD) solid supports. 10 milligrams of purified peptide was coupled to 9 milligrams of PPD and the conjugate was purified twice by column chromatography to remove unconjugated impurities. Rabbits were immunised over a 96 day period. Antibodies were screened and isolated using their binding for the peptides, using routine methods (reviewed in Harlow, E and Lane, D (1988), *Antibodies: a laboratory manual*, CSH Press, NY.).

Antibodies specific for individual subunits were obtained. The antibodies proved to be specific for both mouse and human subunits.

Immunoprecipitation performed under native conditions (50mM Hepes pH8, 90mM KCl, 0.5% TX-100) with these subunit specific antibodies shows that CCT consists of a complex mixture of particles. Different antibodies immunoprecipitate different combinations of subunits (for instance, see Figure 9). Examination of different immunoprecipitations allows the determination of the composition of CCT in any cell type. This information enables the intelligent selection of subunit combinations to be used in *in vitro* construction of CCT, for instance by means of recombinant expression of subunits and their assembly.

Antibody characterisation

2D-PAGE Mouse testis CCT was purified by sucrose gradient fractionation followed by ATP-affinity column chromatography as previously described [14]. Isoelectric focusing (IEF) was carried out according to [45]; and was followed by SDS-PAGE on 8% gels and either silver staining or electrotransfer of proteins to nitro-cellulose, immunoblotting and detection by the ECL system (Amersham) as previously described [14].

Non-denaturing isoelectric focusing

Mouse testis CCT was partially purified by sucrose gradient fractionation according to [14]. The fraction corresponding to 20% sucrose was mixed with an equal volume of sample buffer (40% sucrose w/v, 2% Ampholytes [Resolyte 4 - 8, BDH]) and subjected to non-denaturing isoelectric focusing at 500V for 4 hours. The non-

denaturing IEF was performed in the same way as denaturing IEF described by [45] with the substitution of 40% sucrose for 50% urea and the omission of 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) in the acrylamide gel mixture. Non-denaturing IEF was followed by SDS-PAGE on 8% gels and either silver staining or electrotransfer of proteins to nitro-cellulose, immunoblotting and detection by the ECL system (Amersham).

As shown in Figure 10, mouse testis CCT contains nine subunit species (S1-S() by 2D-PAGE analysis (Fig. 10A and Table 3). *Tcp-1* and seven *Tcp-1* related genes encoding the α, β, γ, δ, ϵ, ζ, η and θ subunits of mouse CCT have been cloned [52]; the testis-expressed subunit, S6, and a co-purifying 63kDa protein (arrowed in Fig. 10A) which may be a further subunit of mouse CCT remain to be cloned. We have obtained DNA sequences of S6 and S7 and have established that they are encoded by two closely related CCTζ genes. We expect that S6 and S7 can substitute for each other in CCT and may confer some tissue specificity in function of CCT.

A previously characterised monoclonal antibody, 91a [41,52], was used to detect CCTα (Fig. 10D). Six antibodies recognising single subunit species, BC-1, GC-1, DC-1, EC-1, TC-1 and THC-2 were produced, which reacted specifically with the β, γ, δ, ϵ, η and θ subunits of CCT respectively (Fig. 10B, 10C, 10E, 10F, 10H and 10I). In addition to CCTζ the antibody ZC-1 also

reacted with the testis-expressed subunit S6 (Fig. 10G), which indicates that these two subunits S6 and S7, have related carboxy terminal sequences and this supports the DNA sequence data that they are encoded by closly related genes. A polyclonal rabbit antibody, UM-1, was made to an amino terminal consensus motif which is highly conserved between all chaperonin sequences and is proposed to be involved in ATP binding and hydrolysis (Table 3) [52,53]. UM-1 reacted with all nine subunits of mouse testis CCT including the testis-expressed subunit S6. CCTβ, CCTϵ, CCTγ, CCTζ and S6 were recognised strongly (Fig. 10J), which CCTα, CCTδ, CCTη and CCTθ were recognised weakly (not shown in Fig. 10J but seen on a longer exposure). In addition to the nine subunits described above UM-1 strongly reacted with the co-purifying 63kDa protein, which suggests that this protein could be an additional subunit species of mouse testis CCT or alternatively a modified isoform of a previously identified CCT subunit.

Resolution of native populations of CCT

Partially purified CCT from mouse testis germ cells resolves into two distinct populations, I and II, when subjected to non-denaturing isoelectric focusing (Fig.11). These populations were further analysed in a second dimension by SDS-PAGE and visualised by silver staining (Fig.11A) or immunoblotting with specific antibodies (Fig. 11B and 11C). A striking difference between the two populations is that a large number of

polypeptides, in addition to subunits of the CCT complex, are associated with II whilst I has only a few co-migrating species (Fig. 11A). It is likely that many of these associated polypeptides are substrates for folding by CCT and that II has a high affinity for binding polypeptide chain substrates and I has a low affinity. To date, only tubulin and actin have been established as substrates for folding by CCT *in vivo* [15,18,16] and both β-tubulin (Fig. 11C) and actin (data not shown) are found associated with II but not with I.

Arrowheads show hsp 70 proteins which resolve away from CCT (Fig. 11A), possibly due to the high pH conditions during sample loading. Small amounts of hsp 70 proteins co-immunoprecipitate with CCT under native conditions [14,52] suggesting that hsp 70 is found associated due to its interaction with polypeptide chains undergoing folding on CCT. Therefore, this technique is a stringent analysis of protein interactions and a reflection of the high affinity of CCT form II for substrates.

Fig. 11B shows CCT populations I and II immunoblotted with a polyclonal rabbit antibody to CCTε, however both populations contain all nine subunits of mouse testis CCT as demonstrated by immunoblotting with the antibodies described above (data not shown). Since resolution of native IEF markers in this system demonstrates that separation is on the basis of charge and not size (data not shown), we propose that I and II represent two

distinct conformations of hetero-oliogmeric CCT which are resolved under these conditions due to different complements of surface charges being exposed in each conformation.

During the biochemical extraction of CCT from mouse testis and from other sources, we often observe limited proteolysis of all the subunits in preparations of intact 900kDa CCT complex. A 30kDa fragment of CCTε (arrowed in Fig. 11B) is detected only in form II. This differential susceptibility to proteolysis indicates that the cleavage site is more accessible in II, which is supporting evidence that I and II adopt different conformations. The size of these proteolytic fragments are between 28 to 30 kDa suggesting that the protease sensitive sites reside in the putative apical polypeptide chain binding region of each CCT subunit [53].

Discussion of antibodies

Antibodies specific for each subunit species will be useful in investigating the combination and arrangement of subunits in the CCT complex and the antibody, UM-1, which recognises all the subunits can be used as a general reagent to identify CCT from other eukaryotes. Characterisation of mouse testis CCT revealed that the testis-expressed subunit, S6 [52], is related to S7 and that a co-purifying 63kDa protein may be a novel subunit of mouse testis CCT. The 63 kDa protein is probably encoded by a novel *Tcp-1*-related gene.

A problem with the biochemical analysis of large

molecular weight complexes such as CCT is in resolving different states even when these states differ by the addition of accessory polypeptides, such as substrates or co-factors. Herein described a non-denaturing isoelectric focusing technique that facilitates the resolution of two forms of CCT. One form is bound to many other polypeptides and is susceptible to proteolysis whilst the other form is bound to only a few other polypeptides and is resistant to proteolysis.

These results suggest that CCT in the cell may adopt two distinct conformations with different affinity for polypeptide chain substrates. The conformation with high binding affinity for substrates may expose a domain which is susceptible to proteolysis in each CCT subunit, and this site may be located within the substrate binding domain since it is in a similar position to the substrate binding domain of GroEL [54,55]. In GroEL, dramatic structural changes have been associated with ATP binding [5,56]. One interpretation of our data is that binding or hydrolysis of ATP acts as a switch between conformational forms of CCT that interact strongly or weakly with polypeptide chain substrates. The wide range of polypeptides found associated with CCT suggests that the physiological substrates for CCT may not be limited to actin and tubulin and implies that CCT has a general role in protein folding in the eukaryotic cytosol.

Production and assembly of polypeptide components of a folding complex

(1) In order to create a complex able to fold a polypeptide, messenger RNA dependent rabbit reticulocyte lysates are used. Synthetic, capped mRNAs encoding each of the CCT subunits to associate are independently synthesised *in vitro* by RNA transcription of each linearised Bluescript™ plasmid containing a full-length coding CCT sequence under the appropriate control sequences. (Sambrook et al, 1989, *Molecular Cloning: a Laboratory Manual*, CSH Press, New York) These RNAs are added to rabbit reticulocyte lysate containing ^{35}S-methionine for synthesis of ^{35}S-labelled CCT subunits, which facilitates detection on native electrophoretic gels or sucrose density gradients. The presence of particular subunits is confirmed also using subunit-specific antibodies, by immunoprecipitation.

Synthesis of subunits together leads to the assembly of a complex whose ability to bind to and/or fold a polypeptide can be tested.

Translating mRNA encoding a substrate polypeptide to be folded, such as actin or tubulin, with the subunits enables detection of complexes of labelled CCT and substrate. Labelled substrate facilitates this. Although rabbit reticulocyte extracts contain endogenous CCT, this does not cycle with the newly synthesised CCT subunits.

(2) The rabbit reticulocyte system of (1) is used to investigate different combinations of subunits to assess their ability to associate and form a complex able to

bind and/or fold a polypeptide of interest. For example, an assembled complex of translation products of mRNAs encoding eight subunits (S1, S2, S3, S4, S5, S6 or S7, S8 and S9) is precipitated using an antibody specific for any one of the subunits (Figures 13 and 14) and its binding to a substrate also translated *in vitro*, such as tubulin, is testable (16).

(3) Subunits are expressed in *E. coli* using appropriate vectors. The combination of subunits chosen for a particular purpose, ie folding of a particular polypeptide, is made in large quantities. The subunits associate to form heteromeric structures of expected sedimentation properties and molecular weight. Subunits may be expressed individually and then purified, with the assembly carried out *in vitro* by mixing CCT subunits in the appropriate combinations.

(4) For large scale production, subunits are expressed using Baculovirus. A nucleotide sequence encoding a subunit is inserted into a suitable expression vector and operably linked to control elements within that vector. The vector and wild-type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine to form a recombinant baculovirus expression vector. The baculovirus vector is packaged into infectious recombinant baculoviruses. Large quantities of subunits are produced which are able to associate to form complexes able to fold a polypeptide substrate.

(5) Folding of a polypeptide substrate, for instance tubulin, actin or luciferase, is tested by a standard technique as described in Frydman et al (17). See also references (4), (15) and (16).

Assembly of mouse CCT from 8 recombinant cDNA clones.

The purpose of these experiments is to show that the 8 cDNA clones encoding the CCTα, β, γ, δ, ϵ, ζ, η and θ subunits are necessary and sufficient to produce recombinant CCT proteins capable of co-assembling into the heteromeric core-CCT complex which we expect from our biochemical and genetic analysis to be constructed from these 8 subunits. We have demonstrated this in the following way. Plasmids containing T3 or T7 promoters upstream of the CCT cDNA open reading frames were linearized with appropriate restriction endonucleases to produce templates for the synthesis of each CCT mRNA by T3 or T7 RNA polymerase. The plasmids which we use are CCTα/pT1611, CCTβ/pTβ2, CCTγ/pTγ7, CCTδ/pTδ2, CCTϵ/pTϵ5, CCTζ/pTζ12, CCTη/pTη29 and CCTθ/pTθ1.

When rabbit reticulocyte lysate was programmed singly with each CCT subunit encoding cDNA a 35S-labelled polypeptide of expected molecular weight was synthesized (Figure 12).

To show assembly, the lysate was programmed with all eight subunit encoding plasmids together. After synthesis the reaction mix was applied to a 10-40% sucrose gradient in order to resolve the assembled CCT complexes from the free subunits as shown previously by

us (Reference 14). Labelled CCT complex can be detected at 20% sucrose density (Figure 13A) and all the subunits are immuno-precipitated using antibodies specific to single subunits (Figure 13B). This demonstrates the principle of co-assembly from the 8 subunit genes. This process is repeatable using subunit proteins expressed in mammalian tissue culture cells, bacteria and yeasts.

Yeast expression

The *Pischia pastoris* heterologous gene expression system has been used to express individual CCT subunits in yeast. Specifically we inserted the CCTε open reading frame into pHIL-D2. This construct was used to transform *Pischia* producing strains expressing CCTε under the control of the AOX1 promoter. These strains, such as WIL5, were shown to produce CCTε in response to induction by methanol, both using small scale cultures (to measure time course of induction) and subsequently by large-scale fermentation. In order to follow expression of mouse CCTε in *Pischia pastoris* in these experiments anti-CCTε antibodies disclosed herein were utilised. Without the antibodies the experiments would not have been possible. All procedures were according to the In Vitrogen manual and as described in Cregg, J.M. et al., 1993 Bio/Technology, Vol. 11, 905-910.

Expression in human cells

CCT subunits were expressed in human cells as free subunits and incorporated into endogenous CCT complex using standard eukaryotic expression vectors such as

pcDNA1. Specifically, the CCTα and CCTβ open reading frames were subcloned into pcDNA1 giving plasmids pEXα and pEXβ. These plasmids were transfected into human 293T cells and CCT protein expression was measured by Western blotting and immunoprecipitation using antibodies disclosed herein. Substantial expression of CCTα and CCTβ was obtained.

Additional Discussion

The symmetry of each ring of the CCT double-torus particle is yet to be determined, although EM-negative stain analysis suggests that it is likely to comprise 8 or 9 subunits [14]. Although the stoichiometric composition of CCT with respect to individual polypeptide species has not yet been determined biochemically, we have confirmed that there are seven species of CCT subunits by 2D-NEPHGE analysis of antibody-affinity purified CCT purified from mouse F9 cells. We have been able to assign eight CCT subunit proteins to eight *Cct* genes including *Tcp-1/Cctα*. The analysis of CCT [14] and the actin chaperonin [15] by negative stain electron microscopy is consistent with the heteromeric nature of the CCT since the particles appear quasi-symmetrical with different sized/shaped subunits. Lewis *et al.* [14] previously showed that mouse testis CCT contained more subunits than human HEp2 cell CCT and Roobol and Carden [19] found differences between brain CCT and testis CCT derived from rats and guinea-pigs. These data are consistent with the idea that a particular CCT particle

is composed of several kinds of subunits, although some subunits may be replaceable depending upon tissue type.

The eight genes (the seven new ones plus the original TCP-1 (alpha)) described herein are sufficient to reconstitute CCT. Certain cell types such as mouse F9 and human HEp-2 contain abundant levels of 7 CCT subunits, with the S2 subunit, encoded by the *CCTe* gene, appearing to be absent, based on protein analysis but it may be rapidly turning over or post-translationally modified causing levels of S2 to be difficult to measure. The role of the mouse testis expressed subunit, S6, is unclear but it probably has an exchangeable function with S7, the product of the *CCTzeta* gene, given that antibodies raised against S7 react with S6 and that we have DNA clones for two closely related CCTζ genes. Northern blotting and protein analysis lead to the conclusion that S1, S2, S3, S4, S5, S7, S8 and S9 are ubiquitously expressed and therefore that they constitute the eight essential components of general CCT activity.

One can build/reconstruct CCT folding machines (complexes) using combinations of these seven components. Specialised proteins may need the presence of an additional, eg tissue specific, subunit or other cofactors to be correctly folded. However, these should be able to be incorporated into CCT machines in combination with the subunits provided herein.

Evolutionary origin of CCT subunit genes

All organisms are classifiable into three primary

kingdoms, eubacteria, archaebacteria and eukaryotes, based upon ribosomal RNA sequences [30]. Several proteins found in archaebacteria and eukaryotic cytosol are similar to each other and those of mitochondria, chloroplasts and eubacteria are also similar to each other [31].

Our sequence analysis has shown that each CCT subunit is very similar to TF55 of an archaebacterium *Sulfolobus shibatae* (Figure 4, Table 2) and only weakly related to GroEL and other chaperonin proteins which are thought to have evolved from a common eubacterial origin [10]. Another archaebacterium, *Pyrodictium occultum*, also has an ATPase complex whose tryptic peptides share 70% identity with TF55 [32]. These results suggest that all the CCT genes evolved from a common ancestor of eukaryotes and archaebacteria. Each eukaryotic *Cct* gene is as divergent from another *Cct* gene as it is from TF55 and this suggests that the eukaryotic genes diverged from one another very early in the eukaryotic lineage. An evolutionary tree of CCT subunits and TF55 based on amino-acid substitutions supports these ideas (Figure 7). If the amino-acid substitution rate of each CCT subunit is constant during evolution [33] and assuming that yeasts and animals diverged 1000-1200 million years ago, then we calculate 1800-2400 million years for the divergence times of CCT subunits.

However, the orthologues of each CCT member in animals, plants and yeasts are much more similar to one

another and this suggests that each subunit has evolved an independent function a long time ago which has been maintained in most eukaryotes. Since CCT is composed of 7-9 subunit species and other chaperonins have only one or two subunit species, it seems likely that the CCT has evolved more complex functions in eukaryotic cytosol. The increased complexity of CCT may have facilitated the evolution of the highly organized eukaryotic cytosol by co-evolution between CCT and substrates.

Function of the CCT

CCT, TF55 and classical chaperonins (GroEL, Hsp60 and Rubisco binding protein) all have ATPase activity. It seems likely that the conservation of sequence in this family reflects the maintenance of ATPase activity rather than the ability to form double torus structures which many unrelated proteins, such as glutamine synthetase [34] can adopt. In GroEL the conformation of the particle changes by rotation of the GroEL subunits relative to the perpendicular axis upon binding ATP [5]. EM analysis of the actin chaperonin in the presence and absence of ATP suggests that similar large conformational changes may occur [15]. There are three highly conserved motifs among the TCP-1, TF55 and GroEL family chaperonins (see Figure 4 in [14]). Two motifs, (I/V)T(N/K)DG(A/V)(T/S) (SEQ. ID NO. 72) and GDGTT(S/T) (SEQ. ID NO. 73), are positioned at the NH_2 terminus and one, V(A/P)GGG (SEQ. ID NO. 74), towards the COOH terminus. These three motifs are also conserved in the

other CCT subunits and interestingly the amino-acid sequences around GDGTT(T/S) (Figure 5) have homologies to the phosphate binding domains for ATP hydrolysis shared by the cAMP-dependent kinase family proteins [25,26]. If these widely separated motifs are indeed involved in ATP binding and hydrolysis, then they may be responsible for the large conformational changes the chaperonins make upon ATP binding. This idea is supported by our knowledge of the structure of cAMP-dependent protein kinase-α [26], Hsp70 [35,36] and actin/myosin [37]; several domains required for ATPase activity are widely separated in the amino-acid chains of these proteins.

One possible reason for the multiplicity of CCT subunit species may be the contributions of each subunit to interaction with other co-factors involved in protein folding processes. We previously showed that CCT binds Hsp70 species [14]. In *E. coli*, DnaK (Hsp70 family) and GroE (Hsp60 family) are thought to cooperate with DnaJ in the folding of newly synthesized proteins [38,39]. It is possible that a similar mechanism involving Hsp70, DnaJ homologues [40] and CCT occurs in eukaryotic cytosol. We found a weak homology between the COOH terminus of CCTδ and Hsp70 family proteins (not shown) and this may suggest a role for CCTδ in interactions between chaperones and co-factors.

It is known that some chaperonins (GroE, Hsp60) are not able to fold substrates which can be folded by CCT; mitochondrial Hsp60 does not fold or form a binary

complex with denatured α-actin [15] and GroEL/ES does not fold luciferase although it can bind it [17]. This suggests that the substrate specificity of classical chaperonins and CCT differs and that CCT might have changed its spectrum of substrate building by evolving a far larger number of subunit species than the single subunit species chaperonins. Sternlicht *et al.* [18] reported that a 40 kDa newly synthesized protein co-purified with CCT in addition to actin and tubulin. This would seem to be a novel substrate of CCT. The general function of CCT, in addition to its folding of actin and tubulin, is illustrated by a genetic experiment in *S. cerevisiae*. Expression of human *HTR3* cDNA which encodes a partial TCP-1 related polypeptide rescued a mutation in amino-acid transport [41]. The HTR3 polypeptide has 96% homology to mouse CCTζ, and is thus the human orthologue of CCTζ. This suggests that CCT might assist folding of a protein affecting amino-acid transport. CCT binds *in vitro* to chromaffin granule membranes coupled to an affinity matrix [42]. This suggests a role for CCT for the regulation of secretory vesicles and membrane trafficking events. A short peptide sequence of the TCP-1-related chaperonin of oats [42] which seems to have evolved the specialized function of folding phytochrome, is more related to CCTδ (Residue nos. 369-384 in Figure 4) than the other CCT subunits. Specialized functions might have been derived from different primordial *Cct* genes in different ancestral organisms.

CCT is believed to act mainly as folding machinery for tubulin and actin by many investigators despite the fact that these proteins show no significant amino-acid sequence homology. Actin and tubulin are extremely abundant proteins in the cell, and thus they were probably the easiest proteins to be detected as substrates [18].

Table 4 summarises proteins known to be folded by CCT. Firefly luciferase has been reported to be folded by CCT *in vitro* (17, 61) and brain CCT was purified as a neurofilament fragment binding protein [19]. Lingappa et al. [62] reported that TCP-1 and/or a TCP-1-like chaperonin is involved in capsid assembly by hepatitis B virus *in vitro*. Thus, CCT may assist in the folding of not only actin and tubulin but also a more wider range of proteins *in vivo*.

Functional analysis of each subunit species will provide further information of the role of CCT *in vivo* in general protein folding events in eukaryotic cells.

We have identified nine subunit species of CCT in mammalian testis and seven major species in mouse F9 cells and we cloned seven *Tcp-1* related genes encoding CCT subunits found in common to testis and F9 cell CCT. *Tcp-1/Cctα* and these genes are so highly conserved from mammals to yeast they can be detected by cross-species Southern blotting analysis. The amino-acid sequences of the mouse and yeast orthologues of two CCT subunit species (CCTα and CCTß) show 61% and 66% sequence

identity, respectively. All CCT subunits contain a motif similar to an ATP binding domain of cAMP-dependent protein kinase and other kinases of this family, suggesting all CCT subunits have independent ATPase activities. The mouse CCT subunits show only 26%-35% homology between one subunit protein species and another and phylogenetic analysis suggests that the divergence time of CCT subunit genes is approximately two times as long as the divergence time of animals and yeasts.

These observations suggest that each subunit has evolved a specific, independent function in addition to the common ATPase function a long time ago and that these functions are maintained in all eukaryotes. The expansion of the number of subunit species compared to other chaperonins may have allowed more complex functions required for folding and assembly of highly evolved eukaryotic proteins.

References

1. Ellis et al., Annu. Rev. Biochem. 1991, **60**: 321-347.
2. Hartl et al., Annu. Rev. Biophys. Biomol. Struct. 1992, **21**: 293-322.
3. Horwich et al., Phil. Trans. R. Soc. Lond. 1993, **21**: 293-322.
4. Langer et al., EMBO J. 1992, **11**: 4757-4765.
5. Saibil et al., Curr. Biol. 1993, **3**: 265-273.
6. Weissman et al., Cell 1994, **78**: 693-702.
7. Georgopoulos et al., J. Mol. Biol. 1973, **76**: 45-60.
8. Horwich et al., Cell 1993, **74**: 909-917.
9. Cheng et al., Nature 1989, **337**: 620-625.
10. Gupta, Biochem. Internatl. 1990, **20**: 833-841.
11. Trent et al., Nature 1991, **354**: 490-493.
12. Hemmingsen, Nature 1992, **357**: 650.
13. Ellis, Nature 1992, **358**: 191-192.
14. Lewis et al., Nature 1992, **358**: 249-252.
15. Gao et al., Cell 1992, **69**: 1043-1050.
16. Yaffe et al., Nature 1992, **358**: 245-248.
17. Frydman et al., EMBO J. 1992, **11**: 4767-4778.
18. Sternlicht et al., Proc. Natl. Acad. Sci. USA 1993, **90**: 9422-9426.
19. Roobol et al., J. Neurochem. 1993, **60**: 2327-2330.
20. Willison et al., Cell 1989, **44**: 727-738.
21. Willison et al., Cell 1989, **57**: 621-632.
22. Waterston et al., Nature Genet. 1992, **1**: 114-123.

23. Sokal et al., W H Freeman & Company 1963.

24. Agard, Science 1993, **260**: 1903-1904.

25. Hanks et al., Science 1988, **241**: 42-52.

26. Taylor et al., Ann. Rev. Cell Biol. 1992, **8**: 429-462.

27. Ursic et al., Mol. Cell Biol. 1991, **11**: 2629-2640.

28. Miklos et al., Proc. Natl. Acad. Sci. USA 1994, in press.

29. Mori et al., Gene 1992, **122**: 381-382.

30. Woese et al., Proc. Natl. Acad. Sci. USA 1990, **87**: 4576-4579.

31. Iwabe et al., Proc. Natl. Acad. Sci. USA 1989, **86**: 9355-9359.

32. Phipps et al., Nature 1993, **361**: 475-477.

33. Kubota et al., Gene 1991, **105**: 269-273.

34. Almassy et al., Nature 1986, **323**: 304-309.

35. Flaherty et al., Proc. Natl. Acad. Sci. USA 1991, **88**: 5041-5045.

36. Bork et al., Proc. Natl. Acad. Sci. USA 1992, **89**: 7290-7294.

37. Rayment et al., Science 1993, **261**: 58-65.

38. Langer et al., Nature 1992, **356**: 683-689.

39. Craig, Science 1993, **260**: 1902-1903.

40. Silver et al., Cell 1993, **74**: 5-6.

41. Segel et al., Proc. Natl. Acad. Sci. USA 1992, **89**: 6060-6064.

42. Creutz et al., J. Biol. Chem. 1994 **269** 32035-32038.

43. O'Farrell, J. Biol. Chem. 1975, **250**: 4007-4021.

44. O'Farrell et al., Cell 1977, **12**: 1133-1142.

45. Corbett et al., Edited by Graham JM, Higgins JA, New Jersey: Humana Press 1992, 219-227.

46. Fernandez et al., Anal Biochem. 1992, **201**: 255-264.

47. Nagata et al., Nucleic Acids Res. 1990, **19**: 2441-2447.

48. Kirchhoff et al., Nucleic Acid Res. 1990, **18**: 4247.

49. Khan et al., Nature Genet 1992, **2**: 180-185.

50. Saitou et al., Mol. Biol. Evol. 1987, **4**: 406-425.

51. Ursic et al., Gene 1988, **68**: 267-274.

52. Kubota et al., Curr. Biol. 1994, **4**: 89-99.

53. Kim et al., TIBS 1994, **19**: 543-548.

54. Fenton et al., Nature 1994, **371**: 614-619.

55. Braig et al., Nature 1994, **371**: 578-586.

56. Chen et al., Nature 1994, **371**: 261-264.

57. Melki et al., J. Cell Biol 1993, **128**: 1301-1310.

58. Gao et al., Mol. Cell. Biol. 1993, **13**: 2478-2485.

59. Rommelaere et al., PNAS USA 1993, **90**: 11975-11979.

60. Gao et al., J. Cell. Biol. 1994, **125**: 989-996.

61. Frydman et al., Nature. 1994, **370**: 111-117.

62. Lingappa et al, J. Cell. Biol 1994, **125**: 99-111.

Table 1. Characteristics of CCT subunits.

Subunit species	Spot on 2D gels	Number of amino acids	Molecular weight
CCTα	S3	556	60513
CCTβ	S4	536	57456
CCTγ	S5	545	60636
CCTδ	S9	539	58073
CCTε	S2	541	59631
CCTζ	S7	531	58011
CCTη	S8	541	59658
CCTθ	S1	548	59562

The spot numbers of CCT subunits on 2D gels (S1-S9) are described in Figure 1. The correspondences of CCT subunit α to spot S3 [14], ε to S2, ζ to S7 and θ to S1 (Figure 10) were determined by Western blotting of 2D gels with antibodies which react specifically with each subunit. The other correspondences are derived from the peptide sequencing data described in Figure 3 and the Western blotting data in Figure 10. Numbers of amino acids and molecular weights of CCT subunits are calculated from the amino-acid sequences in Figure 3 and Figure 8.

Table 2. Homology among mouse CCT subunits and archaebacterial chaperone TF55.

	CCTα	CCTβ	CCTγ	CCTδ	CCTε	CCTζ	CCTη	CCTθ	TF55
CCTα		7(24)	3(13)	4(12)	3(12)	6(24)	5(21)	5(19)	4(16)
CCTβ	35.4		4(8)	4(10)	5(9)	5(14)	4(8)	6(8)	6(8)
CCTγ	32.5	28.4		5(10)	2(2)	5(16)	4(8)	5(11)	2(2)
CCTδ	32.1	28.2	33.1		4(9)	5(10)	3(4)	5(11)	5(10)
CCTε	31.9	31.9	32.2	36.2		5(17)	4(6)	4(6)	3(6)
CCTζ	27.1	25.0	29.2	28.5	31.0		4(8)	3(12)	4(13)
CCTη	34.6	31.9	34.6	28.5	28.4	26.3		4(8)	3(6)
CCTθ	27.5	25.4	27.4	27.8	29.1	22.5	24.0		29.0
TF55	38.7	32.0	34.1	37.1	38.0	35.1	34.9	5(10)	

Percentages of identical amino acids are shown below the diagonal and numbers of gaps/insertions are above the diagonal. Numbers in parentheses are total lengths of gaps/insertions given in amino acid numbers.

Table 3. Immunochemical analysis of CCT subunits

Antibody used for detection	Sequence of peptide immunogen	Corresponding subunit of mouse CCT	Species recognised on 2D gel
THC-2	SGKKDWDDDQND	CCTθ	S1
EC-1	IDDIRKPGESEE	CCTε	S2
91a*	-	CCTα	S3
BC-1	APRKRVPDHHPC	CCTβ	S4
GC-1	NRQTGAPDAGQE	CCTγ	S5
ZC-1	EIMRAGMSSLKG	CCTζ	S6,S7
TC-1	SAGRGRGQARFH	CCTη	S8
DC-1	SILKIDDVVNTR	CCTδ	S9
UM-1	QDDEVGDGTTSVV	chaperonin consensus motif	S1,S2,S3,S4,S5, S6,S7,S8,S9,p63**

91a* denotes a rat monoclonal antibody which recognises the C-terminus of CCTα [2].
p63** denotes a co-purifying 63kDa protein of pI 6.93.

Table 4 Proteins known to be folded or assembled by CCT and co-factors.

Protein	Experiment	Reference
actin	in vitro, in vivo	15,18
actin-RVP	in vitro	57
tubulin	in vitro, in vivo	16,17,18,57,58,59,60
neurofilament(fragment)	in vitro	19
luciferase	in vitro	17,61
hepatitis B virus capsid	in vitro*	62

* TCP-1 or TCP-1-like chaperonin recognized by anti-TCP-1 antibody.

CLAIMS

1. A DNA isolate comprising a nucleotide sequence encoding a subunit of a polypeptide folding complex, the subunit having an amino acid sequence shown in any of SEQ. ID NO.'s 25, 27, 28, 29, 30 and 31 or an amino acid sequence having a significant degree of homology to a shown sequence.

2. A DNA isolate according to claim 1 wherein the nucleotide sequence is any of SEQ. ID NO.'s 55, 57, 58, 59, 60 and 61.

3. A DNA isolate according to claim 1 wherein the nucleotide sequence is an allele, a mutant or derivative, by way of nucleotide deletion, substitution or insertion, of any of the nucleotide sequences shown in SEQ. ID NO.'s 55, 57, 58, 59, 60 and 61.

4. A DNA isolate comprising a nucleotide sequence encoding a subunit of a polypeptide folding complex, the subunit comprising an amino acid sequence which is a mutant or derivative, by way of amino acid deletion, substitution or insertion, of any of the amino acid sequences shown in SEQ. ID NO.'s 25, 27, 28, 29, 30 and 31.

5. A vector comprising DNA according to any one of

claims 1 to 4 and nucleic acid for expression of the encoded subunit.

6. A host cell comprising a vector according to claim 5.

7. A host cell according to claim 6 which is prokaryotic.

8. A host cell according to claim 6 which is eukaryotic.

9. A method of making a subunit of a polypeptide folding complex comprising expression from a vector according to claim 5.

10. A method according to claim 9 wherein the expression results from culturing a host cell comprising the vector.

11. A method according to claim 10 wherein the host cell is prokaryotic.

12. A method according to claim 10 wherein the host cell is eukaryotic.

13. A method of producing a polypeptide folding complex, the method comprising expression of subunits of

the complex from encoding nucleic acid therefor and causing or allowing assembly of the subunits into the complex.

14. A method according to claim 13 wherein the subunits comprise any of the amino acid sequences shown in SEQ. ID NO.'s 25, 26, 27, 28, 29, 30 and 31 or an amino acid sequence having a significant degree of homology to a shown sequence.

15. A method according to claim 14 wherein the nucleic acid encoding any of the subunits has a nucleotide sequence which is an allele, a mutant or derivative, by way of nucleotide deletion, substitution or insertion, of any of the nucleotide sequences shown in SEQ. ID NO.'s 55, 56, 57, 58, 59, 60 and 61.

16. A method according to claim 13 wherein the amino acid sequence of any of the subunits is a mutant or derivative, by way of amino acid deletion, substitution or insertion, of any of the amino acid sequences shown in SEQ. ID NO.'s 25, 26, 27, 28, 29, 30 and 31.

17. A method according to any one of claims 13 to 16 wherein assembly of the subunits into the complex is by *in vitro* mixing following expression.

18. A method according to any one of claims 13 to 16

wherein the subunits are expressed together.

19. A purified polypeptide subunit of a polypeptide folding complex comprising an amino acid sequence shown in any of SEQ. ID NO.'s 25, 27, 28, 29, 30 and 31, or an amino acid sequence having a significant degree of homology to a shown sequence.

20. A polypeptide subunit of a polypeptide folding complex comprising an amino acid sequence which is a mutant or derivative, by way of amino acid deletion, substitution or insertion, of any of the amino acid sequences shown in SEQ. ID NO.'s 25, 27, 28, 29, 30 and 31.

21. A polypeptide folding complex comprising polypeptide subunits selected from subunits according to claim 19 or claim 20, optionally a subunit comprising the amino acid sequence which is shown in SEQ. ID NO. 26 or an amino acid sequence which is a mutant or derivative, by way of amino acid deletion, substitution or insertion, of the amino acid sequence shown in SEQ. ID NO. 26, or an amino acid sequence having a significant degree of homology to the sequence shown in SEQ. ID NO. 26, and optionally a subunit comprising the amino acid sequence which is shown in SEQ. ID NO. 24 or an amino acid sequence which is a mutant or derivative, by way of amino acid deletion, substitution or

insertion, of the amino acid sequence shown in SEQ. ID NO. 24, or an amino acid sequence having a significant degree of homology to the sequence shown in SEQ. ID NO. 24.

22. Use of a polypeptide folding complex according to claim 21 in the folding of a polypeptide.

23. A method of folding a polypeptide employing a polypeptide folding complex according to claim 21.

24. A method according to claim 23 wherein the folding of the polypeptide follows assembly of the polypeptide folding complex from subunits thereof, following production of the subunits by recombinant expression.

25. An antibody or fragment thereof able to bind specifically to a subunit of a polypeptide folding complex other than TCP-1 (CCTα - SEQ. ID NO. 24).

26. An antibody or fragment thereof according to claim 25 wherein the subunit to which the antibody or fragment thereof is able to bind is any of CCTβ (SEQ. ID NO. 25), CCTγ (SEQ. ID NO. 26), CCTδ (SEQ. ID NO. 27), CCTϵ (SEQ. ID NO. 28), CCTζ (SEQ. ID NO. 29), CCTη (SEQ. ID NO. 30) and CCTθ (SEQ. ID NO. 31).

27. An antibody or fragment thereof according to claim

26 having binding specificity for a peptide which has an amino acid sequence shown in any of SEQ. ID NO.'s 1, 2, 3, 4, 5, 6 and 7, or which has an amino acid sequence which is a variant of any of those shown in SEQ. ID NO.'s 1, 2, 3, 4, 5, 6 and 7.

28. A method of obtaining an antibody or fragment thereof according to claim 25 comprising administering a peptide which has an amino acid sequence shown in any of SEQ. ID NO.'s 1, 2, 3, 4, 5, 6 and 7, or which has an amino acid sequence which is a variant of any of those shown in SEQ. ID NO.'s 1, 2, 3, 4, 5, 6 and 7, to a mammal and recovering antibody from the mammal.

Fig.1a.

Fig.1b.

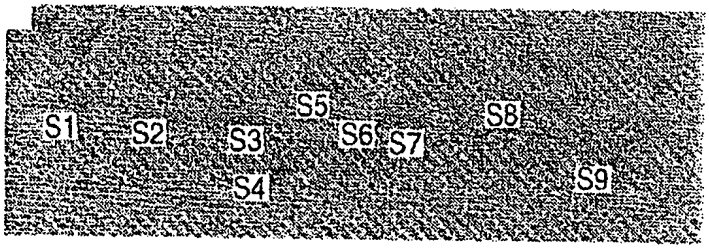


Fig.1c.

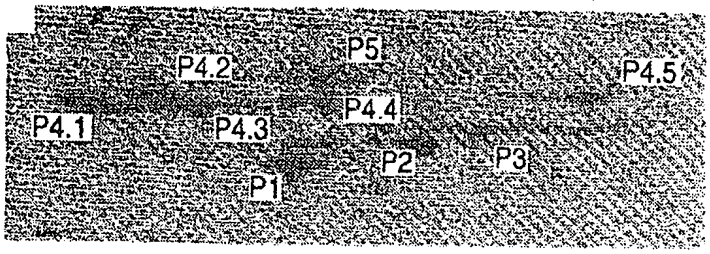


Fig.1d.

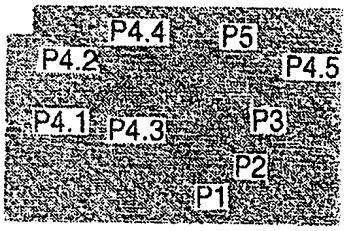


Fig.1e.

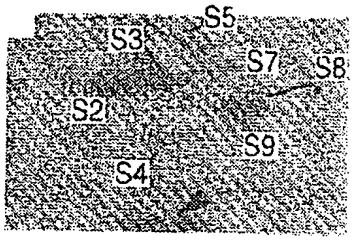


Fig.1f.

Mouse			Bovine		
Spot	pI	M.W. (KD)	Spot	pI	M.W. (KD)
S1	6.25	62	P4.1	6.1	60
S2	6.35	61	P4.2	6.3	61
S3	6.5	60	P4.3	6.45	60
S4	6.65	53	P1	6.8	52
S5	6.85	65	P5	6.85	65
S6	6.9	62.5	P4.4	6.85	60.5
S7	6.95	62	P2	7.0	56
S8	7.1	57	P3	7.1	58
S9	>7.2	56	P4.5	>7.2	61

Fig.2.

pTε5 (M)	KSRLMGLEALKSHIMAAKAVANTMRTSL**GP**NGLD**K**MMVDKDGDV	Group 1
p384 (H)	KSRLMGLEALKSHIMAAKAVANTMRTSL**GP**NGLD**K**MMVDKDGDV	
pG4 (C)	QKRITGVEAVKSHILAARAVANTLRTSL**GP**RGLD**K**MLVSPDGDV	
pTδ2 (M)	QDRDKPAQIRFSNISAAKAVADAIRTSL**GP**KGMD**K**MIQDGKGDV	Group 2
pG3 (C)	KDKDKPESVRNSNIVAAKAVADAVRTSL**GP**RGMD**K**MIQSGNGDV	
Tcp-1 (M)	GDRSTGEAVRSQNVMAAASIANIVKSSF**GP**VGLD**K**MLVDDIGDV	Group 3
TCP1 (H)	GDRSTGETIRSQNVMAAASIANIVKSSL**GP**VGLD**K**MLVDDIGDV	
pG1 (C)	GKRTTGQGIRSQNVTAAVAIANIVKSSL**GP**VGLD**K**MLVDDVGDV	
TCP1 (S)	GDRQSGQDVRTQNVMACQAVSNIVKTSL**GP**VGLD**K**MLVDDIGDV	
pCBL80 (M)	TDSSQGIPQLVSNISACQVIAEAVRTTL**GP**RGMD**K**LIVDGRGKA	Group 4
pTγ7 (M)	TKRESGRKVQSGNINAAKTIADIIRTCL**GP**KSMM**K**MLLDPMGGI	Group 5
pTζ12 (M)	AEVARVQAALAVNISAARGLQDVLRTNL**GP**KGTM**K**MLVSGAGDI	Group 6
pG2 (C)	AELARHAAALELNISGARGLQDVMRSNL**GP**KGTL**K**MLVSGAGDI	
pTβ2 (M)	ADEERAEIARLSSFIGAIAIGDLVKSTL**GP**KAMD**K**ILLSSGRDA	Group 7
p383 (H)	ADEERAETARLTSFIGAIAIGDLVKSTL**GP**KGMD**K**ILLSSGRDA	
4950 (S)	VTEERAENARLSAFVGAIAVGDLVKSTL**GP**KGMD**K**LLQSASSNT	

Fig.3a.

```
MASLSLAPVNIFKAGADEERAEIARLSSFIGAIAIGDLVKSTLGPKGMDK   050
ILLSSGRDAALMVTNDGATILKNIGVDNPAAKVLVDMSRVQDDEVGDGTT   100
                                              P1(T31)
SVTVLAAELLREAESLIAKKIHPQTIISGWREATKAAREALLSSAVDHGS   150
DEARFWQDLMNIAGTTLSSKLLTHHKDHFTKLAVEAGLRLKGSGNLEAIH   200
VIKKLGGSLADSYLDEGFLLDKKIGVNQPKRIENAKILIANTGMDTDKIK   250
                                  X-X-  B4(R162)
IFGSRVRVDSTAKVAEIEHAEKEKMKEKVERILKHGINCFINRQLIYNYP   300
EQLFGAAGVMAIEHADFAGVERLALVTGGEIASTFDHPELVKLGSCKLIE   350
EVMIGEDKLIHFSGVALGEACTIVLRGATQQILDEAERSLHDALCVLAQT   400
VKDPRTVYGGGCSEMLMAHAVTQLANRTPGKEAVAMESFAKALRMLPTII   450
ADNAGYDSADLVAQLRAAHSEGHITAGLDMKEGTIGDMAVLGITESFQVK   500
RQVLLSAAEAAEVILRVDNIIKAAPRKRVPDHHPC   535
```

Fig.3b.

```
MMGHRPVLVLSQNTKRESGRKVQSGNINAAKTIADIIRTCLGPKSMMKML   050
LDPMGGIVMTNDGNAILREIQVQHPAAKSMIEISRTQDEEVGDGTTSVII   100
LAGEMLSVAEHFLEQQMHPTVVISAYRMALDDMISTLKKISTPVDVNNRE   150
MMLSIINSSITTKVISRWSSLACNIALDAVKTVQFEENGRKEIDIKKYAR   200
VEKIPGGIIEDSCVLRGVMINKDVTHPRMRRYIKNPRIVLLDSSLEYKKG   250
ESQTDIEITREEDFTRILQMEEEYIHQLCEDIIQLKPDVVITEKGISDLA   300
                N-------QQ------XX---- P5(T36)
QHYLMRANVTAIRRVRKTDNNRIARACGARIVSRPEELREDDVGTGAGLL   350
EIKKIGDEYFTFITDCKDPKACTILLRGASKEILSEVERNLQDAMQVCRN   400
VLLDPQLVPGGGASEMAVAHALTEKSKAMTGVEQWPYRAVAQALEVIPRT   450
      P5(T32)                 B1(R133)
LIQNCGASTIRLLTSLRAKHTQESCETWGVNGETGTLVDMKELGIWEPLA   500
VKLQTYKTAVETAVLLLRIDDIVSGHKKKGDDQNRQTGAPDAGQE   545
```

Fig.3c.

```
MPENVASRSGAPTAGPGSRGKSAYQDRDKPAQIRFSNISAAKAVADAIRT  050
SLGPKGMDKMIQDGKGDVTITNDGATILKQMQVLHPAARMLVELSKAQDI  100
EAGDGTTSVVIIAGSLLDSCTKLLQKGIHPTIISESFQKALEKGLEILTD  150
          X    P3(T31)
MSRPVQLSDRETLLNSATTSLNSKVVSQYSSLLSPMSVNAVMKVIDPATA  200
TSVDLRDIKIVKKLGGTIDDCELVEGLVLTQKVANSGITRVEKAKIGLIQ  250
FCLSAPKTDMDNQIVVSDYAQMDRVLREERAYILNLVKQIKKTGCNVLLI  300
QKSILRDALSDLALHFLNKMKIMVVKDVEREDIEFICKTIGTKPVAHIDQ  350
FTADMLGSAELAEEVSLNGSGKLFKITGCTSPGKTVTIVVRGSNKLVIEE  400
AERSIHDALCVIRCLVKKRALIAGGGAPEIELALRLTEYSRTLSGMESYC  450
VRAFADAMEVIPSTLAENAGLNPISTVTELRNRHAQGEKTTGINVRKGGI  500
          P3(T50)                 X
SNILEEMVVQPLLVSVSALTLATETVRSILKIDDVVNTR  539
```

Fig.3d.

```
MASVGTLAFDEYGRPFLIIKDQDRKSRLMGLEALKSHIMAAKAVANTMRT  050
SLGPNGLDKMMVDKDGDVTITNDGATILSMMDVDHQIAKLMVELSKSQDD  100
EIGDGTTGVVVLAGALLEEAEQLLDRGIHPIRIADGYEQAARIAIQHLDK  150
ISDKVLVDINNPEPLIQTAKTTLGSKVINSCHRQMAEIAVNAVLTVADME  200
RRDVDFELIKVEGKVGGRLEDTKLIKGVIVDKDFSHPQMPKKVVDAKIAI  250
LTCPFEPPKPKTKHKLDVMSVEDYKALQKYEKEKFEEMIKQIKETGANLA  300
ICQWGFDDEANHLLLQNGLPAVRWVGGPEIELIAIATGGRIVPRFSELTS  350
EKLGFAGVVQEISFGTTKDKMLVIEKCKNSRAVTIFIRGGNKMIIEEAKR  400
SLHDALCVIRNLIRDNRVVYGGGAAEISCALAVSQEADKCPTLEQYAMRA  450
FADALEVIPMALSENSGMNPIQTMTEVRARQVKESNPALGIDCLHKGSND  500
MQYQHVIETLIGKKQQISLATQMVRMILKIDDIRKPGESEE  541
```

Fig.3e.

```
MAAVKTLNPKAEVARAQAALAVNISAARGLQDVLRTNLGPKGTMKMLVSG   050
AGDIKLTKDGNVLLHEMQIQHPTASLIAKVATAQDDITGDGTTSNVLIIG   100
ELLKQADLYISEGLHPRIITEGFEAAKEKALQFLEQVKVSKEMDRETLID   150
VARTSLRTKVHAELADVLTEAVVDSILAIRKKDEPIDLFMVEIMEMKHKS   200
ETDTSLIRGLVLDHGARHPDMKKRVENAYILTCNVSLEYEKTEVNSGFFY   250
KSAEEREKLVKAERKFIEDRVKKIIELKKKVCGDSDKGFVVINQKGIDPF   300
                  B3(R170)
SLDALAKEGIVALRRAKRRNMERLTLACGGIALNSFDDLNPDCLGHAGLV   350
YEYTLGEEKFTFIEKCNNPRSVTLLVKGPNKHTLTQIKDAIRDGLRAVKN   400
AIDDGCVVPGAGAVEVALAEALIKYKPSVKGRAQLGVQAFADALLIIPKV   450
LAQNSGFDLQETLVKVQAEHSESGQLVGVDLSTGEPMVAAEMGVWDNYCV   500
KKOLLHSCTVIATNILLVDEIMRAGMSSLKG   531
```

Fig.3f.

```
MMPTPVILLKEGTDSSQGIPQLVSNISACQVIAEAVRTTLGPRGMDKLIV   050
DGRGKATISNDGATILKLLDVVHPAAKTLVDIAKSQDAEVGDGTTSVTLL   100
                                                  P2
AAEFLKQVKPYVEEGLHPQIIIRAFRTATQLAVNKIKEIAVTVKKQDKVE   150
    X      B2(R115)          B2(R124)
QRKMLEKCAMTALSSKLISQQKVFFAKMVVDAVMMLDELLQLKMIGIKKV   200
QGGALEESQLVAGVAFKKTFSYAGFEMQPKKYKNPKIALLNVELELKAEK   250
DNAEIRVHTVEDYQAIVDAEWNILYDKLEKIHQSGAKVILSKLPIGDVAT   300
QYFADRDMFCAGRVPEEDLKRTMMACGGSIQTSVNALVPDVLGHCQVFEE   350
TQIGGERYNFFTGCPKAKTCTIILRGGAEQFMEETERSLHDAIMIVRRAI   400
KNDSVVAGGGAIEMELSKYLRDYSRTIPGKQQLLIGAYAKALEIIPRQLC   450
DNAGFDATNILNKLRARHAQGGMWYGVDINNENIADNFQAFVWEPAMVRI   500
NALTAASEAACLIVSVDETIKNPRSTVDPPAPSAGRGRGQARFH   544
```

Fig.4.

```
                      20                  40                  60                  80                 100
CCTα  ----------------MEGPLSVFGDRSTGEAVRSQNVMAAASIANIVKSSFGPVGLDKMLVDDIGD--VTITNDGATILKLLEVEHPAAKVLCELADLQ
CCTβ  ---------MASLSLAPVNIFKAGADEERAEIARLSSFIGAIAIGDLVKSTLGPKGMDKILLSSGRDAALMVTNDGATILKNIGVDNPAAKVLVDMSRVQ
CCTγ  -----------MMGHRPVLVLSQNTKRESGRKVQSGNINAAKTIADIIRTCLGPKSMMKMLLDPMGG--IVMTNDGNAILREIQVQHPAAKSMIEISRTQ
CCTδ  MPENVASRSGAPTAGPGSRGKSAYQDRDKPAQIRFSNISAAKAVADAIRTSLGPKGMDKMIQDGKGD--VTITNDGATILKQMQVLHPAARMLVELSKAQ
CCTε  MASVGTLAFDEYGRPFLIIKDQDRKSRLMGLEALKSHIMAAKAVANTMRTSLGPNGLDKMMVDKDGD--VTITNDGATILSMMDVDHQIAKLMVELSKSQ
CCTζ  --------------MAAVKTLNPKAEVARAQAALAVNISAARGLQDVLRTNLGPKGTMKMLVSGAGD--IKLTKDGNVLLHEMQIQHPTASLIAKVATAQ
CCTη  ------------MMPTPVILLKEGTDSSQGIPQLVSNISACQVIAEAVRTTLGPRGMDKLIVDGRGK--ATISNDGATILKLLDVVHPAAKTLVDIAKSQ
CCTθ  -----MALHVPKAPGFAQMLKDGAKHFSGLEEAVYRNIQACKELAQTTRTAYGPNGMNKMVINRLEK--LFVTNDAATILRELEVQHPAAKMIVMASHMQ
TF55  ---MATATVATTPEGIPVIILKEGSSRTYGKEALRANIAAVKAIEEALKSTYGPRGMDKMFVDSLGD--ITITNDGATILDKMDLQHPTGKLLVQIAKGQ
Cons. ................PVIILS.G.DRS.GEEAL.SNISAAKAIADAVRTSLGPKGMDKMLVDG.GD..VTITNDGATILKELQVQHPAAKLLVELAK.Q
                          K           R               KST               I          MD                IS

                     120                 140                 160                 180                 200
CCTα  DKEVGDGTTSVVIIAAELLKNADELVKQKIHPTSVISGYRLACKEAVRYINENLIINTDELGR---DCLINTAKTSMSSKIIGINGDYFANMVVDAVLAV
CCTβ  DDEVGDGTTSVTVLAAELLREAESLIAKKIHPQTIISGWREATKAAREALLSSAVDHGSDEAR-FWQDLMNIAGTTLSSKLLTHHKDHFTKLAVEAGLRL
CCTγ  DEEVGDGTTSVIILAGEMLSVAEHFLEQQMHPTVVISAYRMALDDMISTLKKISTPVDVNNRE----MMLSIINSSITTKVISRWSSLACNIALDAVKTV
CCTδ  DIEAGDGTTSVVIIAGSLLDSCTKLLQKGIHPTIISESFQKALEKGLEILTDMSRPVQLSDRE----TLLNSATTSLNSKVVSQYSSLLSPMSVNAVMKV
CCTε  DDEIGDGTTGVVVLAGALLEEAEQLLDRGIHPIRIADGYEQAARIAIQHLDKISDKVLVDINN--PEPLIQTAKTTLGSKVINSCHRQMAEIAVNAVLTV
CCTζ  DDITGDGTTSNVLIIGELLKQADLYISEGLHPRIITEGFEAAKEKALQFLEQVKVSKEMDRET-----LIDVARTSLRTKVHAELADVLTEAVVDSILAI
CCTη  DAEVGDGTTSVTLLAAEFLKQVKPYVEEGLHPQIIIRAFRTATQLAVNKIKEIAVTVKKQDKVEQRKMLEKCAMTALSSKLISQQKVFFAKMVVDAVMML
CCTθ  EQEVGDGTNFVLVFAGALLELAEELLRIGLSVSEVISGYEIACKKAHEILPELVCCSAKNLRD--VDEVSSLLRTSIMSKQYGSETFLAKLIAQACVSIF
TF55  DEETADGTKTAVILAGELAKKAEDLLYKEIHPTIIVSGYKKAEEIALKTIQDIAQPVSINDTD----VLRKVALTSLGSKAVAGAREYLADLVVKAVAQV
Cons. DDEVGDGTTSVVILAGELLK.AE.LL.KGIHPTIIISGYR.A.EKAL.TL.EIA.PV..DDR.....MLIN.A.TSLSSKVIS...DLLA.MVVDAVL.V
                 V A                      V   FE   K    I   S   N                               F  IA

                     220                 240                 260                 280                 300
CCTα  KYTDARGQPRYPVNSVNILKAHGRSQIESMLINGYALNCVVGSQG---MPKRIVNAKIACLDFSLQKTKMKL-GVQVVITDPEKLDQIRQRESDITKERI
CCTβ  -------KGSGNLEAIHVIKKLGGSLADSYLDEGFLLDKKIGVNQ----PKRIENAKILIANTGMDTDKIKIFGSRVRVDSTAKVAEIEHAEKEKMKEKV
CCTγ  QFEENGRKEIDIKKYARVEKIPGGIIEDSCVLRGVMINKDV--------THPRMRRYIKNPRIVLLDSSLEY-KKGESQTDIEITREEDFTRILQMEEEY
CCTδ  --IDPATATSVDLRDIKIVKKLGGTIDDCELVEGLVLTQKVANSG----ITRVEKAKIGLIQFCLSAPKTDM-DNQIVVSDYAQMDRVLREERAYILNLV
CCTε  ---ADMERRDVDFELIKVEGKVGGRLEDTKLIKGVIVDKDFSHPQ---MPKKVVDAKIAILTCPFEPPKPKT-KHKLDVMSVEDYKALQKYEKEKFEEMI
CCTζ  ----RKKDEPIDLFMVEIMEMKHKSETDTSLIRGLVLDHGARHPD---MKKRVENAYILTCNVSLEYEKTEV-NSGFFYKSAEEREKLVKAERKFIEDRV
CCTη  -------DELLQLKMIGIKKVQGGALEESQLVAGVAFKKTFSYAGFEMQPKKYKNPKIALLNVELELKAEKD-NAEIRVHTVEDYQAIVDAEWNILYDKL
CCTθ  -----PDSGNFNVDNIRVCKILGSGIYSSSVLHGMVFKKETEGD-----VTSVKDAKIAVYSCPFDGMITET-KGTVLIKTAEELMNFSKGEENLMDAQV
TF55  -AELRGDKWYVDLDNVQIVKKHGGSINDTQLVYGIVVDKEVVHPG---MPKRIENAKIALLDASLEVEKPEL-DAEIRINDPTQMHKFLEEEENILKEKV
Cons. .......KE.VDL..I.I.KKLGGSI.DS.LV.GVVLDK.V.HPG...MPKRIENAKIALLN.SLE..K.EL.K..IRV.D.E....I.KAE..IMEEKV
                     V                  I                 V                 K    V I                 K
```

```
              .        320         .         340         .         360         .         380         .       400
CCTα   QKILATGANVILTT---------GGIDDMYLKYFVEAGAMAVRRVLKRDLKHVAKASGASILSTLANLEGEETFEVTMLGQAEEVVQERICDDELILIKN
CCTβ   ERILKHGINCFINR---------QLIYNYPEQLFGAAGVMAIEHADFAGVERLALVTGGEIASTFDHPELVK---LGSCKLIEEVMIGEDK---LIHFSG
CCTγ   IHQLCEDIIQLKPD----VVITEKGISDLAQHYLMRANVTAIRRVRKTDNNRIARACGARIVSRPEELREDD---VGTGAGLLEIKKIGDE--YFTFITD
CCTδ   KQIKKTGCNVLLIQ----KSILRDALSDLALHFLNKMKIMVVKDVEREDIEFICKTIGTKPVAHIDQFTADM---LGSAELAEEVSLNGSG-KLFKITGC
CCTε   KQIKETGANLAICQ---------WGFDDEANHLLLQNGLPAVRWVGGPEIELIAIATGGRIVPRFSELTSEK---LGFAGVVQEISFGTTK-DKMLVIEK
CCTζ   KKIIELKKKVCGDSDKGFVVINQKGIDPFSLDALAKEGIVALRRAKRRNMERLTLACGGIALNSFDDLNPDC---LGHAGLVYEYTLGEEK---FTFIEK
CCTθ   KAIAGTGANVIVTG---------GKVADIALHYANKYNIMLVRLNSKWDLRRLCKTVGATALPKLTPPVQEE---MGHCDSVYLSEVGDTQ--VVVFKHE
CCTη   EKIHQSGAKVILSK---------LPIGDVATQYFADRDMFCAGRVPEEDLKRTMMACGGSIQTSVNALVPDV---LGHCQVFEETQIGGER---YNFFTG
TF55   DKIAATGANVVICQ---------KGIDEVAQHYLAKKGILAVRRAKKSDLEKLARATGGRVISNIDELTSQD---LGYAALVEERKVGEDK---MVFVEG
Cons.  KKIL.TGANVIL.Q......I..KGIDD.ALHYLAKAGIMAVRRV.K.DLERIAKACGGRIVS.FDELT.D....LGHAGLVEEV..GEDK...F.FIEG
                                                                L    A        E     C     G

              .        420         .         440         .         460         .         480         .       500
CCTα   TKARTSASIILRGANDFMCDEMERSLHDALCVVKRVLELKSVVPGGGAVEAALSIYLENYATNMGSREQLAIAEFARSLLVIPNTLAVNAAQDSTDLVAK
CCTβ   VALGEACTIVLRGATQQILDEAERSLHDALCVLAQTVKDPRTVYGGGCSEMLMAHAVTQLANRTPGKEAVAMESFAKALRMLPTIIADNAGYDSADLVAQ
CCTγ   CKDPKACTILLRGASKEILSEVERNLQDAMQVCRNVLLDPQLVPGGGASEMAVAHALTEKSKAMTGVEQWPYRAVAQALEVIPRTLIQNCGASTIRLLTS
CCTδ   TSPGKTVTIVVRGSNKLVIEEAERSIHDALCVIRCLVKKRALIAGGGAPEIELALRLTEYSRTLSGMESYCVRAFADAMEVIPSTLAENAGLNPISTVTE
CCTε   CKNSRAVTIFIRGGNKMIIEEAKRSLHDALCVIRNLIRDNRVVYGGGAAEISCALAVSQEADKCPTLEQYAMRAFADALEVIPMALSENSGMNPIQTMTE
CCTζ   CNNPRSVTLLVKGPNKHTLTQIKDAIRDGLRAVKNAIDDGCVVPGAGAVEVALAEALIKYKPSVKGRAQLGVQAFADALLIIPKVLAQNSGFDLQETLVK
CCTη   CPKAKTCTIILRGGAEQFMEETERSLHDAIMIVRRAIKNDSVVAGGGAIEMELSKYLRDYSRTIPGKQQLLIGAYAKALEIIPRQLCDNAGFDATNILNK
CCTθ   KEDGAISTIVLRGSTDNLMDDIERAVDDGVNTFKVLTRDKRLVPGGGATEIELAKQITSYGETCPGLEQYAIKKFAEAFEAIPRALAENSGVKANEVISK
TF55   AKNPKSVSILIRGGLERVVDETERALRDALGTVADVIRDGRAVAGGGAVEIEIAKRLRKYAPQVGGKEQLAIEAYANAIEGLIMILAENAGLDPIDKLMQ
Cons.  CKNPKAVTILLRGANK.ILDEAERSLHDALCVVRNVIKD.RVVAGGGAVEIELAKALT.YA.T.PGKEQLAIRAFADALEVIPRTLAENAG.DPIDTVTK
            G SC     G     E            K  L      P      M               Y                      S     LL

              .        520         .         540         .         560         .         580         .
CCTα   LRAFHNEAQVNPERKNLKWIGLDLVHGKPRDNKQAGVFEPTIVKVKSLKFATEAAITILRIDDLIKLHPESKDDKHGSYENAVHSGALDD
CCTβ   LRAAHSE--------GHITAGLDMKEGTIGDMAVLGITESFQVKRQVLLSAAEAAEVILRVDNIIKAAPRKRVPDHHPC-----------
CCTγ   LRAKHTQ-------ESCETWGVNGETGTLVDMKELGIWEPLAVKLQTYKTAVETAVLLLRIDDIVSGHKKKGDDQNRQTGAPDAGQE---
CCTδ   LRNRHAQ--------GEKTTGINVRKGGISNILEEMVVQPLLVSVSALTLATETVRSILKIDDVVNTR-------------------
CCTε   VRARQVK-------ESNPALGIDCLHKGSNDMQYQHVIETLIGKKQQISLATQMVRMILKIDDIRKPGESEE-----------------
CCTζ   VQAEHSE--------SGQLVGVDLSTGEPMVAAEMGVWDNYCVKKQLLHSCTVIATNILLVDEIMRAGMSSLKG---------------
CCTη   LRARHAQ--------GGMWYGVDINNENIADNFQAFVWEPAMVRINALTAASEAACLIVSVDETIKNPRSTVDPPAPSAGRGRGQARFH-
CCTθ   LYSVHQE------GNKNVGLDIEAEVPAVKDMLEASILDTYLGKYWAIKLATNAAVTVLRVDQIIMAKPAGGPKPPSGKKDWDDDQND--
TF55   LRSLHEN-------ETNKWYGLNLFTGNPEDMWKLGVIEPALVKMNAIKAATEAVTLVLRIDDIVAAGKKGGSEPGGKKEKEEKSSED--
Cons.  LRA.H.E.......EGNKT.GVDL.TG.P.DM.ELGVWEPLLVK.QALKLATEAA.LILRIDDIVKAGPS.GD................
                        W  LN          A I          I           V   I
                            I
```

Fig.4 (Cont).

Fig.5.

CCTα	KEVGDGTTSVVIIA
CCTβ	DEVGDGTTSVTVLA
CCTγ	EEVGDGTTSVIILA
CCTδ	IEAGDGTTSVVIIA
CCTε	DEIGDGTTGVVVLA
CCTζ	DITGDGTTSNVLII
CCTη	AEVGDGTTSVTLLA
CCTθ	QEVGDGTNFVLVFA
cAPK-α	KTLGTGSFGRVMLV
PKC-α	MVLGKGSFGKVMLA
CaMII-α	EELGKGAFSVVRRC
SNF1	KTLGEGSFGKVKLA
cdc2+	EKIGEGTYGVVYKA
CDC7	DKIGEGTFSSVYKA
Raf	TRIGSGSFGTVYKG
Src	VKLGQGCFGEVWMG
Abl	HKLGGGQYGEVYEG
EGFR	KVLGSGAFGTVYKG
INSR	RELGQGSFGMVYEG
PDGFR	RTLGSGAFGQVVEA

Fig.7.

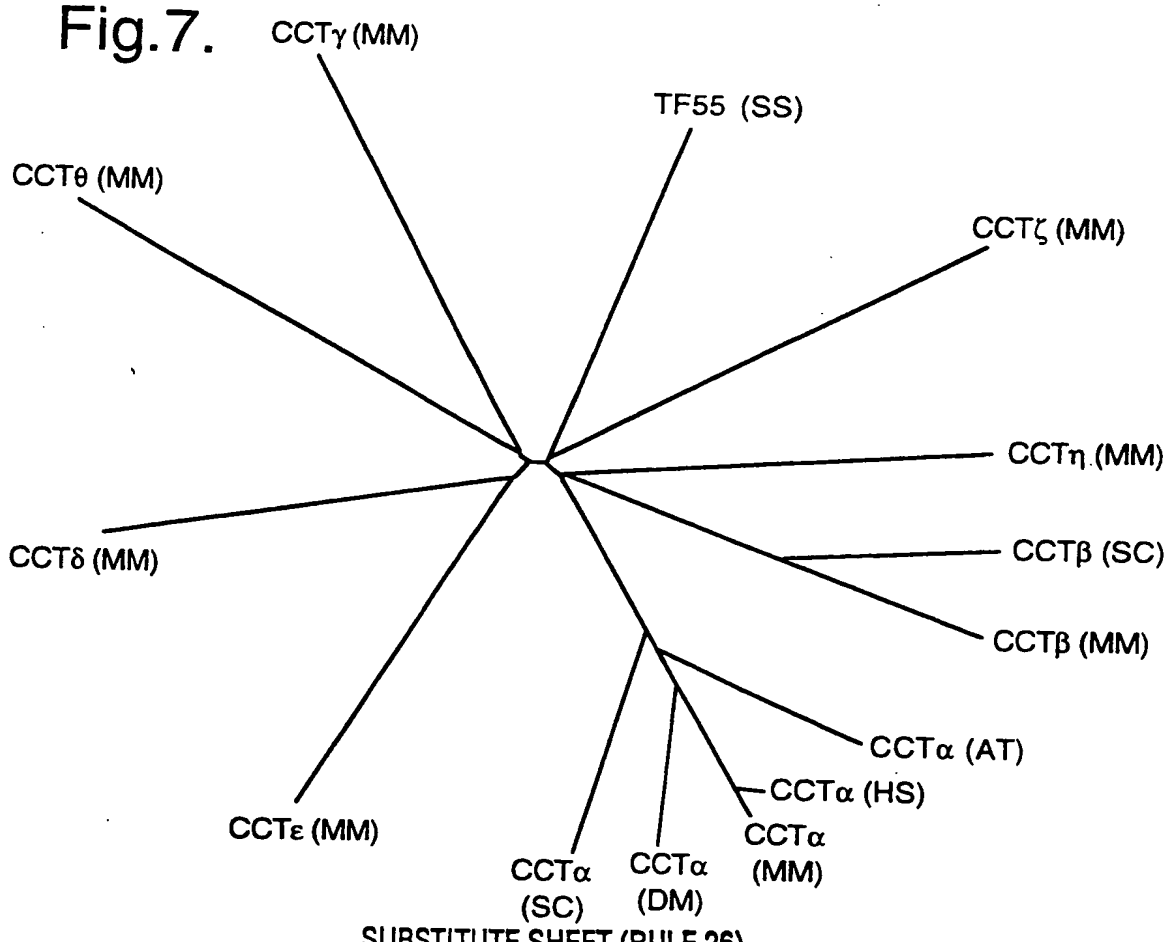


Fig.6a.

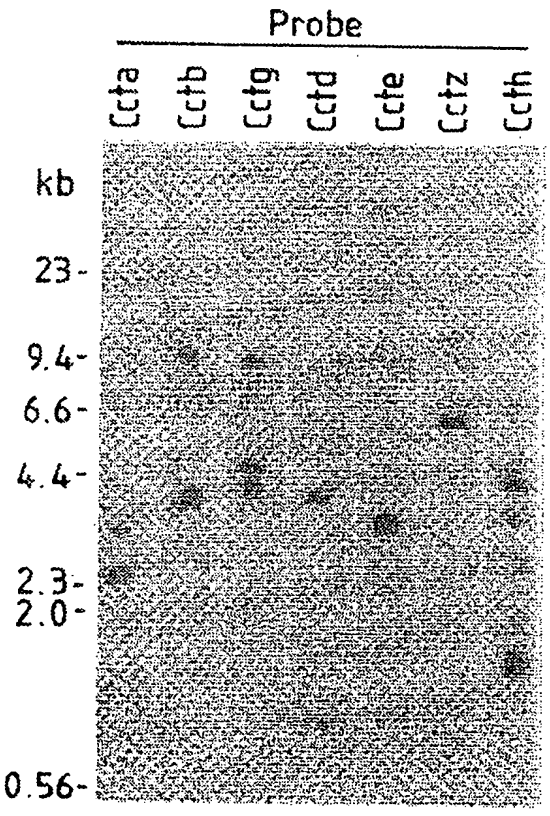


Fig.6b.

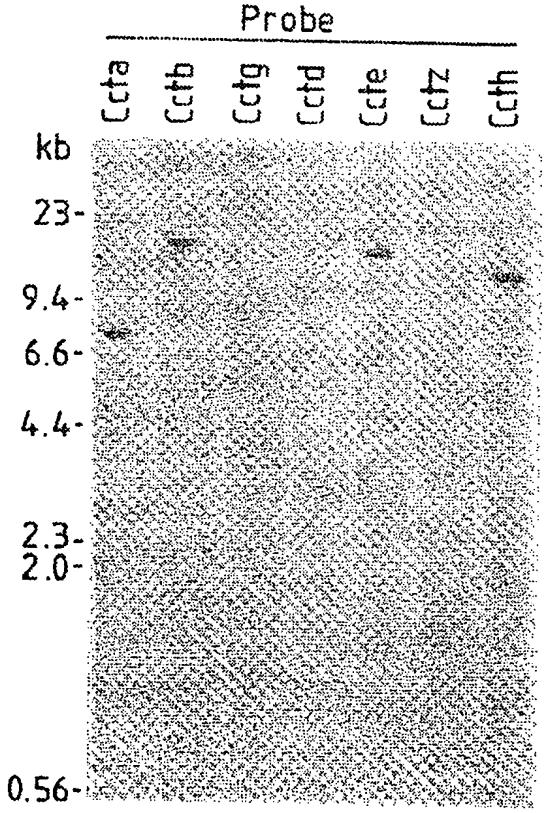


Fig.8a.

(Ccta)

Tcp-1 cDNA sequence

```
TCTCAGAGCGGAGCTATCGGTTGTCTGCGGCCGTAGTGAGCTTTGCTTCGTTTCCTGAAGATGGAGGGTCCTTTGTCCGTGTTCGGGGAC
                                                    M  E  G  P  L  S  V  F  G  D

CGCAGCACTGGGGAGGCGGTCCGCTCCCAGAATGTTATGGCTGCAGCTTCTATTGCCAACATTGTTAAAAGTTCTTTTGGGCCAGTTGGC
R  S  T  G  E  A  V  R  S  Q  N  V  M  A  A  A  S  I  A  N  I  V  K  S  S  F  G  P  V  G

TTGGATAAAATGTTGGTGGATGATATTGGTGATGTAACCATTACTAACGATGGTGCCACCATCCTGAAGTTACTGGAGGTAGAACATCCC
L  D  K  M  L  V  D  D  I  G  D  V  T  I  T  N  D  G  A  T  I  L  K  L  L  E  V  E  H  P

GCAGCCAAAGTTCTGTGTGAGCTGGCTGACCTGCAAGACAAAGAAGTTGGAGATGGAACTACCTCAGTGGTAATCATTGCAGCGGAGCTT
A  A  K  V  L  C  E  L  A  D  L  Q  D  K  E  V  G  D  G  T  T  S  V  V  I  I  A  A  E  L

CTGAAAAATGCAGATGAGCTAGTCAAACAGAAAATTCATCCAACATCAGTTATTAGTGGCTATCGTCTTGCCTGCAAGGAAGCGGTGCGT
L  K  N  A  D  E  L  V  K  Q  K  I  H  P  T  S  V  I  S  G  Y  R  L  A  C  K  E  A  V  R

TATATCAATGAGAACCTGATTATTAACACAGACGAACTTGGAAGAGACTGTCTGATCAATACTGCTAAGACATCCATGTCTTCCAAAATT
Y  I  N  E  N  L  I  I  N  T  D  E  L  G  R  D  C  L  I  N  T  A  K  T  S  M  S  S  K  I

ATTGGAATAAATGGTGATTACTTTGCTAATATGGTAGTAGATGCTGTGCTTGCTGTTAAATACACAGATGCCAGAGGCCAGCCTCGCTAT
I  G  I  N  G  D  Y  F  A  N  M  V  V  D  A  V  L  A  V  K  Y  T  D  A  R  G  Q  P  R  Y

CCAGTCAATTCTGTTAATATTCTGAAAGCCCATGGGAGAAGTCAGATAGAAAGCATGCTGATCAATGGCTATGCGCTCAATTGTGTGGTT
P  V  N  S  V  N  I  L  K  A  H  G  R  S  Q  I  E  S  M  L  I  N  G  Y  A  L  N  C  V  V

GGATCTCAGGGCATGCCCAAGAGAATAGTTAATGCAAAAATTGCTTGTCTTGACTTCAGCCTGCAGAAAACAAAAATGAAGCTTGGTGTA
G  S  Q  G  M  P  K  R  I  V  N  A  K  I  A  C  L  D  F  S  L  Q  K  T  K  M  K  L  G  V

CAGGTGGTTATTACAGACCCTGAGAAATTGGACCAAATTAGACAGAGAGAATCGGATATCACCAAGGAGAGAATTCAGAAGATCCTGGCA
Q  V  V  I  T  D  P  E  K  L  D  Q  I  R  Q  R  E  S  D  I  T  K  E  R  I  Q  K  I  L  A

ACTGGTGCCAATGTTATTCTAACCACTGGTGGCATTGATGATATGTATCTCAAGTATTTTGTGGAAGCTGGTGCCATGGCTGTTAGGAGA
T  G  A  N  V  I  L  T  T  G  G  I  D  D  M  Y  L  K  Y  F  V  E  A  G  A  M  A  V  R  R

GTTTTAAAACGAGACCTGAAGCATGTTGCAAAAGCTTCTGGAGCAAGTATCCTGTCTACGCTGGCCAATTTGGAAGGCGAAGAAACTTTT
V  L  K  R  D  L  K  H  V  A  K  A  S  G  A  S  I  L  S  T  L  A  N  L  E  G  E  E  T  F

GAAGTGACGATGTTGGGACAAGCGGAAGAGGTCGTACAGGAGAGAATTTGTGATGATGAGCTGATCTTAATCAAAAATACTAAGGCTCGT
E  V  T  M  L  G  Q  A  E  E  V  V  Q  E  R  I  C  D  D  E  L  I  L  I  K  N  T  K  A  R

ACATCTGCTTCAATCATCTTACGAGGAGCAAATGATTTCATGTGTGATGAAATGGAGCGCTCTTTACATGATGCTCTTTGTGTGGTGAAG
T  S  A  S  I  I  L  R  G  A  N  D  F  M  C  D  E  M  E  R  S  L  H  D  A  L  C  V  V  K

AGAGTTTTGGAGTTGAAATCTGTGGTCCCAGGTGGAGGTGCTGTAGAAGCTGCCCTGTCCATATACCTGGAAAACTATGCAACAAACATG
R  V  L  E  L  K  S  V  V  P  G  G  G  A  V  E  A  A  L  S  I  Y  L  E  N  Y  A  T  N  M

GGATCTCGGGAACAGCTTGCTATTGCAGAGTTTGCAAGATCTCTGCTTGTGATTCCTAATACACTGGCAGTGAATGCTGCCCAGGACTCC
G  S  R  E  Q  L  A  I  A  E  F  A  R  S  L  L  V  I  P  N  T  L  A  V  N  A  A  Q  D  S

ACCGACCTGGTTGCCAAGTTAAGAGCTTTTCACAATGAGGCTCAAGTGAACCCGGAACGTAAAAATCTAAAGTGGATTGGTCTTGATTTG
T  D  L  V  A  K  L  R  A  F  H  N  E  A  Q  V  N  P  E  R  K  N  L  K  W  I  G  L  D  L

GTCCATGGGAAACCACGAGACAACAAGCAAGCAGGGGTGTTTGAACCAACCATAGTTAAAGTGAAGAGCCTGAAGTTCGCAACAGAGGCT
V  H  G  K  P  R  D  N  K  Q  A  G  V  F  E  P  T  I  V  K  V  K  S  L  K  F  A  T  E  A

GCAATCACCATCCTTCGGATTGACGATCTGATAAAATTACACCCAGAAAGCAAAGACGATAAACACGGAAGTTATGAAAATGCTGTTCAC
A  I  T  I  L  R  I  D  D  L  I  K  L  H  P  E  S  K  D  D  K  H  G  S  Y  E  N  A  V  H

TCTGGAGCCCTTGATGACTGATTGGATTTCCCTCTTATTTATAACAGTGTCAGGTGCAATGCCGTAGCCTTGGGTGTCTCACATTAAAGT
S  G  A  L  D  D  *

ACAGCAAGCTGT          poly(A)*
```

Fig.8b.

(Cctb)

pTbeta2 cDNA Sequence

```
GGCTTCCCGGCGCGGATTGTGAGGGGTTCTCTCCTCTACGCGGAATCCGTCGGAACCATGGCTTCCCTTTCCCTCGCACCTGTTAATATC
                                                    M  A  S  L  S  L  A  P  V  N  I

TTCAAGGCTGGAGCTGATGAAGAGAGGGCCGAGATAGCTCGCCTGTCGTCGTTTATCGGTGCCATCGCCATTGGAGACTTGGTGAAGAGC
F  K  A  G  A  D  E  E  R  A  E  I  A  R  L  S  S  F  I  G  A  I  A  I  G  D  L  V  K  S

ACTTTGGGACCGAAGGGCATGGACAAAATTCTCCTAAGCAGTGGACGAGACGCCGCTCTGATGGTGACCAACGACGGTGCTACCATTCTC
T  L  G  P  K  G  M  D  K  I  L  L  S  S  G  R  D  A  A  L  M  V  T  N  D  G  A  T  I  L

AAGAACATTGGTGTGGACAACCCCGCAGCAAAGGTTCTAGTTGATATGTCAAGGGTTCAAGATGATGAAGTTGGTGATGGCACTACCTCT
K  N  I  G  V  D  N  P  A  A  K  V  L  V  D  M  S  R  V  Q  D  D  E  V  G  D  G  T  T  S

GTTACTGTCTTAGCCGCAGAGCTGCTCCGGGAAGCAGAATCTTTAATTGCAAAAAAGATACATCCACAGACCATCATCTCAGGTTGGAGA
V  T  V  L  A  A  E  L  L  R  E  A  E  S  L  I  A  K  K  I  H  P  Q  T  I  I  S  G  W  R

GAAGCCACAAAGGCAGCAAGAGAGGCCCTGCTGAGCTCCGCTGTGGATCATGGTTCTGATGAAGCCAGATTCTGGCAGGACTTAATGAAC
E  A  T  K  A  A  R  E  A  L  L  S  S  A  V  D  H  G  S  D  E  A  R  F  W  Q  D  L  M  N

ATTGCAGGAACGACATTGTCCTCAAAGCTCCTTACTCACCACAAGGACCACTTTACTAAACTGGCCGTGGAAGCGGGTCTCAGACTGAAA
I  A  G  T  T  L  S  S  K  L  L  T  H  H  K  D  H  F  T  K  L  A  V  E  A  G  L  R  L  K

GGCTCTGGCAACCTGGAGGCGATTCATGTCATCAAGAAACTAGGTGGGAGTCTGGCAGACTCCTATCTAGATGAAGGTTTTCTTTTGGAT
G  S  G  N  L  E  A  I  H  V  I  K  K  L  G  G  S  L  A  D  S  Y  L  D  E  G  F  L  L  D

AAAAAAATTGGAGTAAATCAACCAAAGAGAATTGAAAATGCTAAAATTCTTATTGCAAATACTGGGATGGATACAGACAAAATAAAGATA
K  K  I  G  V  N  Q  P  K  R  I  E  N  A  K  I  L  I  A  N  T  G  M  D  T  D  K  I  K  I

TTTGGCTCTCGGGTAAGAGTTGATTCCACAGCAAAGGTTGCAGAGATAGAACATGCAGAAAAGGAGAAGATGAAGGAGAAAGTTGAACGT
F  G  S  R  V  R  V  D  S  T  A  K  V  A  E  I  E  H  A  E  K  E  K  M  K  E  K  V  E  R

ATTCTTAAGCATGGAATAAATTGCTTTATTAACAGACAGTTAATTTATAATTACCCTGAACAACTCTTCGGCGCTGCTGGCGTCATGGCT
I  L  K  H  G  I  N  C  F  I  N  R  Q  L  I  Y  N  Y  P  E  Q  L  F  G  A  A  G  V  M  A

ATTGAGCATGCGGATTTCGCAGGTGTGGAGCGCCTCGCTCTTGTCACAGGTGGTGAGATTGCCTCTACCTTTGATCACCCAGAACTTGTG
I  E  H  A  D  F  A  G  V  E  R  L  A  L  V  T  G  G  E  I  A  S  T  F  D  H  P  E  L  V

AAGCTTGGAAGTTGCAAACTTATTGAAGAAGTTATGATCGGGGAAGATAAACTCATTCACTTTTCTGGGGGTTGCCCTTGGTGAGGCATGC
K  L  G  S  C  K  L  I  E  E  V  M  I  G  E  D  K  L  I  H  F  S  G  V  A  L  G  E  A  C

ACCATTGTGCTTCGTGGTGCCACTCAGCAAATTCTGGATGAAGCTGAACGATCTCTGCATGATGCTCTTTGTGTTCTTGCTCAGACTGTA
T  I  V  L  R  G  A  T  Q  Q  I  L  D  E  A  E  R  S  L  H  D  A  L  C  V  L  A  Q  T  V

AAAGATCCTAGAACAGTTTACGGGGGAGGCTGCTCTGAGATGCTGATGGCCCATGCTGTGACACAGCTTGCCAACAGAACCCCAGGAAAA
K  D  P  R  T  V  Y  G  G  G  C  S  E  M  L  M  A  H  A  V  T  Q  L  A  N  R  T  P  G  K

GAAGCTGTAGCAATGGAGTCGTTTGCAAAAGCCCTGAGAATGTTGCCGACCATCATAGCCGACAATGCGGGCTATGACAGTGCAGATCTG
E  A  V  A  M  E  S  F  A  K  A  L  R  M  L  P  T  I  I  A  D  N  A  G  Y  D  S  A  D  L

GTGGCACAGCTCCGAGCTGCTCACAGTGAAGGCCATATAACTGCTGGACTGGATATGAAGGAAGGTACCATCGGCGATATGGCAGTACTG
V  A  Q  L  R  A  A  H  S  E  G  H  I  T  A  G  L  D  M  K  E  G  T  I  G  D  M  A  V  L

GGTATAACAGAGAGTTTTCAAGTGAAGCGACAGGTTCTTCTGAGTGCGGCTGAAGCAGCAGAGGTGATCCTGCGAGTGGACAACATTATC
G  I  T  E  S  F  Q  V  K  R  Q  V  L  L  S  A  A  E  A  A  E  V  I  L  R  V  D  N  I  I

AAAGCAGCACCAAGGAAACGTGTCCCCGATCACCACCCCTGTTAAGCATTCCCATTTGCTGATGAACTCTGGGCCAGTTCATAGCAAAGT
K  A  A  P  R  K  R  V  P  D  H  H  P  C  .

TGTACTTGGAAGACTTCAACCTTTAAAGAAGACTGGTGGAATTGACCTGTCCATGATAGCCTTAAGTTTGAACATTAGCTGACCTTCTGT

GTTAAACATGGGTCTAATTTATTTACTGTTTCATTTTCCATACAATTCAGTTGATTTACAAGTTCATTTCTCATACTGTGTATTAAAATA

AAAATCCAGTTACTTAGCCCTTAAAAAAAAAAAA
```

Fig.8c.

(Cctg)

pTgamma7 cDNA Sequence

```
TGTCATTCCTGAGGAGGCAGCGTTTTCTTCGCTGCTCTTCTCTCCAGAAGGTTCTGCCGATTCCCCCCAGCTCTGGAGAGTCGGCTCTG

CGTCGTGCCGCCATGATGGGCCACCGTCCAGTGCTCGTGCTCAGTCAGAACACAAAGCGTGAATCTGGAAGAAAAGTTCAGTCTGGAAAT
            M  M  G  H  R  P  V  L  V  L  S  Q  N  T  K  R  E  S  G  R  K  V  Q  S  G  N

ATCAATGCTGCAAAGACAATTGCAGACATCATCCGGACCTGCTTGGGACCTAAATCTATGATGAAGATGCTTTTGGACCCAATGGGAGGC
I  N  A  A  K  T  I  A  D  I  I  R  T  C  L  G  P  K  S  M  M  K  M  L  L  D  P  M  G  G

ATCGTGATGACCAATGATGGCAATGCCATTCTTCGAGAGATTCAAGTCCAGCATCCCGCAGCAAAGTCCATGATTGAAATTAGCAGGACC
I  V  M  T  N  D  G  N  A  I  L  R  E  I  Q  V  Q  H  P  A  A  K  S  M  I  E  I  S  R  T

CAGGATGAAGAGGTTGGAGATGGGACCACATCAGTAATTATTCTTGCGGGAGAAATGCTCTCTGTGGCTGAACACTTCCTAGAGCAGCAG
Q  D  E  E  V  G  D  G  T  T  S  V  I  I  L  A  G  E  M  L  S  V  A  E  H  F  L  E  Q  Q

ATGCACCCAACAGTGGTGATCAGTGCTTACCGCATGGCACTGGATGATATGATCAGCACTCTGAAGAAAATCAGTACTCCTGTTGATGTC
M  H  P  T  V  V  I  S  A  Y  R  M  A  L  D  D  M  I  S  T  L  K  K  I  S  T  P  V  D  V

AATAACCGTGAGATGATGTTGAGCATCATCAATAGCTCTATTACTACAAAAGTCATCAGTCGGTGGTCCTCCTTGGCATGCAACATTGCA
N  N  R  E  M  M  L  S  I  I  N  S  S  I  T  T  K  V  I  S  R  W  S  S  L  A  C  N  I  A

CTGGATGCTGTTAAGACTGTGCAGTTTGAAGAGAATGGCAGAAAGGAAATTGACATCAAAAAATATGCAAGGGTAGAAAAGATACCCGGG
L  D  A  V  K  T  V  Q  F  E  E  N  G  R  K  E  I  D  I  K  K  Y  A  R  V  E  K  I  P  G

GGCATCATTGAAGACTCATGTGTCTTACGTGGAGTTATGATTAACAAGGATGTGACCCATCCAAGGATGCGCCGCTATATTAAGAATCCT
G  I  I  E  D  S  C  V  L  R  G  V  M  I  N  K  D  V  T  H  P  R  M  R  R  Y  I  K  N  P

CGAATTGTGCTATTGGATTCTTCTTTGGAGTACAAGAAAGGAGAAAGCCAGACCGACATCGAGATTACCCGGGAGGAGGACTTCACGCGG
R  I  V  L  L  D  S  S  L  E  Y  K  K  G  E  S  Q  T  D  I  E  I  T  R  E  E  D  F  T  R

ATCCTGCAGATGGAGGAGGAGTACATCCATCAGCTGTGTGAGGACATCATCCAGCTGAAGCCTGACGTGGTCATCACAGAGAAGGGCATC
I  L  Q  M  E  E  E  Y  I  H  Q  L  C  E  D  I  I  Q  L  K  P  D  V  V  I  T  E  K  G  I

TCAGATTTAGCTCAGCACTACCTCATGCGGGCCAATGTCACAGCCATTCGTAGAGTCCGGAAAACAGACAATAATCGCATTGCTAGAGCC
S  D  L  A  Q  H  Y  L  M  R  A  N  V  T  A  I  R  R  V  R  K  T  D  N  N  R  I  A  R  A

TGTGGGGCACGGATAGTCAGCCGACCTGAGGAACTGAGAGAAGATGATGTTGGTACAGGGGCAGGCTTATTGGAAATCAAGAAGATTGGA
C  G  A  R  I  V  S  R  P  E  E  L  R  E  D  D  V  G  T  G  A  G  L  L  E  I  K  K  I  G

GATGAGTACTTTACATTCATCACTGACTGCAAAGACCCAAAGGCCTGCACCATTCTTCTTAGAGGAGCCAGCAAAGAGATACTCTCGGAA
D  E  Y  F  T  F  I  T  D  C  K  D  P  K  A  C  T  I  L  L  R  G  A  S  K  E  I  L  S  E

GTAGAACGCAACCTCCAGGATGCCATGCAAGTGTGCCGCAATGTTCTACTGGACCCTCAGCTGGTGCCTGGTGGTGGAGCCTCTGAGATG
V  E  R  N  L  Q  D  A  M  Q  V  C  R  N  V  L  L  D  P  Q  L  V  P  G  G  G  A  S  E  M

GCTGTGGCTCATGCCTTGACAGAAAAATCTAAGGCCATGACTGGTGTGGAACAATGGCCATATAGAGCTGTGGCCCAGGCGTTAGAGGTC
A  V  A  H  A  L  T  E  K  S  K  A  M  T  G  V  E  Q  W  P  Y  R  A  V  A  Q  A  L  E  V

ATCCCTCGGACCTTGATCCAGAATTGTGGGGCCAGTACCATTCGTCTGCTTACCTCCCTTCGGGCCAAGCACACACAGGAGAGTTGTGAG
I  P  R  T  L  I  Q  N  C  G  A  S  T  I  R  L  L  T  S  L  R  A  K  H  T  Q  E  S  C  E

ACCTGGGGTGTGAATGGTGAGACTGGTACCTTGGTGGACATGAAAGAGCTGGGTATTTGGGAGCCATTGGCTGTGAAGCTACAAACGTAC
T  W  G  V  N  G  E  T  G  T  L  V  D  M  K  E  L  G  I  W  E  P  L  A  V  K  L  Q  T  Y

AAAACAGCAGTAGAGACTGCAGTTCTGCTTCTGCGGATTGATGACATTGTCTCTGGCCACAAGAAGAAGGGTGATGACCAGAACCGGCAA
K  T  A  V  E  T  A  V  L  L  L  R  I  D  D  I  V  S  G  H  K  K  K  G  D  D  Q  N  R  Q

ACTGGTGCTCCAGATGCTGGCCAGGAGTGAGTGCTGAGCACGGTGACTTCCATGCACAGAACCAGCAGTCTCCCCTTTCCTGAGCCAGAG
T  G  A  P  D  A  G  Q  E  .

TTCCAGGAACACTGTGGACATCTTTGTTTGCGAAGGATCAGGTTGAGGGGCAGCCCCCAGTCTGTCCCATCTCAGTTTGCAAAAAGCACT

GACACGTATCTCTTCTATTGTAAGCTTTCCATTTAGTTTGCTTCCAATGATTAAATCTAAGTCATTTGAAAAAAAAAAAAA
```

Fig.8d.

(Cctd)

pTdelta2 cDNA Sequence

```
GGCTCTCCAGCAGGCCCCTTGCTCGGCTTCCGCCTTCTCTCCCGGCGTCGCTTTCTGGAAGGTTCGTGGAGGAGGCGGTCAGGGAGACC

GTTACTCCACAGCAAGCCGGAATCCGTGTCCATCCGTCCTCCTGAACCCGCGCAGACGCCACCAAGGTCGCCATGCCGGAGAACGTAGCT
                                                                  M  P  E  N  V  A

TCCCGAAGCGGGGCGCCCACCGCCGGGCCCGGCAGCCGCGGGAAAAGCGCCTACCAGGACCGCGACAAGCCAGCCCAGATCCGCTTCAGC
S  R  S  G  A  P  T  A  G  P  G  S  R  G  K  S  A  Y  Q  D  R  D  K  P  A  Q  I  R  F  S

AATATTTCCGCGGCCAAAGCGGTTGCTGATGCTATTAGAACAAGCCTTGGACCTAAAGGAATGGACAAAATGATTCAAGATGGAAAAGGC
N  I  S  A  A  K  A  V  A  D  A  I  R  T  S  L  G  P  K  G  M  D  K  M  I  Q  D  G  K  G

GATGTGACCATTACAAATGATGGTGCCACCATTCTGAAACAAATGCAGGTATTGCATCCAGCAGCCAGAATGCTGGTGGAATTGTCTAAA
D  V  T  I  T  N  D  G  A  T  I  L  K  Q  M  Q  V  L  H  P  A  A  R  M  L  V  E  L  S  K

GCTCAAGACATAGAAGCAGGAGATGGCACCACGTCGGTTGTCATCATTGCTGGCTCTCTTTTAGACTCCTGTACCAAACTTCTGCAGAAA
A  Q  D  I  E  A  G  D  G  T  T  S  V  V  I  I  A  G  S  L  L  D  S  C  T  K  L  L  Q  K

GGTATACATCCAACCATCATTTCCGAGTCATTCCAGAAAGCTTTGGAAAAGGGTCTTGAAATCCTTACTGACATGTCTCGACCTGTGCAA
G  I  H  P  T  I  I  S  E  S  F  Q  K  A  L  E  K  G  L  E  I  L  T  D  M  S  R  P  V  Q

CTGAGCGACAGAGAAACTTTGTTAAATAGCGCAACTACTTCATTGAATTCAAAGGTTGTCTCTCAGTATTCAAGTCTACTCTCTCCAATG
L  S  D  R  E  T  L  L  N  S  A  T  T  S  L  N  S  K  V  V  S  Q  Y  S  S  L  L  S  P  M

AGTGTCAATGCGGTGATGAAAGTGATTGACCCAGCCACAGCTACCAGTGTAGATCTTCGAGATATTAAAATAGTTAAGAAGCTTGGTGGG
S  V  N  A  V  M  K  V  I  D  P  A  T  A  T  S  V  D  L  R  D  I  K  I  V  K  K  L  G  G

ACAATAGATGACTGTGAGCTGGTGGAAGGCCTCGTTCTCACACAGAAAGTAGCAAATTCTGGCATAACAAGAGTTGAAAAGGCTAAGATT
T  I  D  D  C  E  L  V  E  G  L  V  L  T  Q  K  V  A  N  S  G  I  T  R  V  E  K  A  K  I

GGGCTTATTCAGTTTTGCTTATCTGCTCCTAAAACAGATATGGATAATCAAATAGTAGTATCTGACTATGCCCAGATGGATCGAGTGCTT
G  L  I  Q  F  C  L  S  A  P  K  T  D  M  D  N  Q  I  V  V  S  D  Y  A  Q  M  D  R  V  L

CGAGAGGAGAGAGCCTATATTTTAAATTTGGTGAAGCAAATTAAGAAAACAGGATGTAATGTCCTTCTCATACAGAAGTCTATCCTGAGA
R  E  E  R  A  Y  I  L  N  L  V  K  Q  I  K  K  T  G  C  N  V  L  L  I  Q  K  S  I  L  R

GATGCCCTTAGTGATCTTGCATTACATTTTCTGAATAAGATGAAGATTATGGTGGTTAAGGACGTTGAAAGAGAAGACATTGAATTCATC
D  A  L  S  D  L  A  L  H  F  L  N  K  M  K  I  M  V  V  K  D  V  E  R  E  D  I  E  F  I

TGTAAGACAATTGGAACCAAACCAGTTGCTCACATTGACCAGTTCACTGCTGACATGCTGGGTTCTGCTGAGTTAGCAGAGGAAGTCAGT
C  K  T  I  G  T  K  P  V  A  H  I  D  Q  F  T  A  D  M  L  G  S  A  E  L  A  E  E  V  S

TTAAATGGTTCTGGAAAACTATTCAAGATTACAGGTTGTACAAGCCCAGGGAAAACAGTTACAATTGTCGTACGTGGTTCTAACAAACTG
L  N  G  S  G  K  L  F  K  I  T  G  C  T  S  P  G  K  T  V  T  I  V  V  R  G  S  N  K  L

GTGATTGAAGAAGCTGAGCGCTCCATTCATGATGCTCTCTGTGTCATCCGATGCTTAGTAAAGAAAAGAGCTCTTATTGCAGGAGGTGGT
V  I  E  E  A  E  R  S  I  H  D  A  L  C  V  I  R  C  L  V  K  K  R  A  L  I  A  G  G  G

GCTCCAGAAATAGAGCTGGCCCTCAGACTGACAGAGTACTCCCGAACACTGAGTGGTATGGAGTCCTACTGTGTTCGTGCTTTCGCGGAT
A  P  E  I  E  L  A  L  R  L  T  E  Y  S  R  T  L  S  G  M  E  S  Y  C  V  R  A  F  A  D

GCTATGGAAGTCATTCCATCTACACTAGCTGAAAATGCTGGCCTGAATCCCATTTCTACAGTAACAGAGCTAAGAAACCGCCATGCCCAA
A  M  E  V  I  P  S  T  L  A  E  N  A  G  L  N  P  I  S  T  V  T  E  L  R  N  R  H  A  Q

GGAGAAAAAACTACAGGCATTAATGTCCGAAAGGGTGGGATCTCCAACATTTTGGAGGAAATGGTTGTTCAGCCTCTGTTGGTGTCAGTC
G  E  K  T  T  G  I  N  V  R  K  G  G  I  S  N  I  L  E  E  M  V  V  Q  P  L  L  V  S  V

AGTGCTTTGACCTTAGCAACTGAAACTGTGCGGAGCATTCTGAAAATCGATGATGTGGTAAATACTCGATAATCTGGATAAAAGGATGGT
S  A  L  T  L  A  T  E  T  V  R  S  I  L  K  I  D  D  V  V  N  T  R  .

TGACTGCATCATTATGGACAGAAGTACTGTGGCTGGAATGAAGGACAACCACCTTGTTCCTTGTCTGGAAGCTTCAGAGTTTTTGGACAT

TGTCTTCCAGTTGGCATTTGTTTGAAATTGTATTGAAACAATTTAATGAAAACATTAAATACTTGGTTTTAAACTCCAAAAAAAAAAAAA

AAAAAAA
```

Fig.8e.
(Ccte)

pTepsilon5 cDNA Sequence

```
GGATTCTGCGTCCTCTCGCGAGAGGGAGGTGCGGTGTGTCTCTCTTCCCGAGTGGTCCCCGCTGGTTACTGGAGGAGCGCTCGTTCTAGT

TCGTCCACCATGGCGTCCGTGGGGACCCTCGCCTTCGATGAGTATGGGCGCCCCTTTCTCATTATCAAGGACCAAGATCGCAAGTCCCGT
        M  A  S  V  G  T  L  A  F  D  E  Y  G  R  P  F  L  I  I  K  D  Q  D  R  K  S  R

CTCATGGGGCTTGAGGCCCTCAAGTCTCACATCATGGCTGCCAAAGCTGTAGCAAACACAATGCGGACGTCACTGGGACCAAACGGGCTG
L  M  G  L  E  A  L  K  S  H  I  M  A  A  K  A  V  A  N  T  M  R  T  S  L  G  P  N  G  L

GACAAGATGATGGTTGATAAGGATGGCGATGTGACTATAACAAACGATGGTGCCACCATTCTAAGCATGATGGATGTTGATCATCAGATT
D  K  M  M  V  D  K  D  G  D  V  T  I  T  N  D  G  A  T  I  L  S  M  M  D  V  D  H  Q  I

GCCAAGCTGATGGTTGAACTGTCCAAATCCCAGGATGATGAAATTGGAGATGGGACCACAGGAGTGGTTGTCTTGGCTGGGGCCTTGTTG
A  K  L  M  V  E  L  S  K  S  Q  D  D  E  I  G  D  G  T  T  G  V  V  V  L  A  G  A  L  L

GAAGAAGCTGAACAGCTGCTGGACCGAGGCATTCACCCAATCAGAATTGCTGATGGCTACGAGCAGGCTGCCCGAATTGCAATACAACAC
E  E  A  E  Q  L  L  D  R  G  I  H  P  I  R  I  A  D  G  Y  E  Q  A  A  R  I  A  I  Q  H

CTGGACAAGATCAGCGATAAAGTGCTTGTCGACATAAACAACCCTGAACCTCTGATTCAGACTGCAAAAACCACGCTGGGCTCCAAAGTG
L  D  K  I  S  D  K  V  L  V  D  I  N  N  P  E  P  L  I  Q  T  A  K  T  T  L  G  S  K  V

ATTAACAGCTGTCACCGACAGATGGCTGAGATCGCCGTGAATGCCGTCCTCACGGTGGCAGATATGGAGCGGAGAGATGTTGACTTTGAG
I  N  S  C  H  R  Q  M  A  E  I  A  V  N  A  V  L  T  V  A  D  M  E  R  R  D  V  D  F  E

CTCATTAAAGTGGAAGGCAAAGTAGGTGGGCGTCTGGAAGACACCAAGCTCATAAAGGGTGTGATCGTCGACAAGGACTTCAGCCACCCA
L  I  K  V  E  G  K  V  G  G  R  L  E  D  T  K  L  I  K  G  V  I  V  D  K  D  F  S  H  P

CAGATGCCGAAAAAAGTGGTAGATGCTAAGATTGCGATTCTCACGTGTCCATTTGAGCCACCTAAACCTAAGACAAAGCACAAGCTGGAT
Q  M  P  K  K  V  V  D  A  K  I  A  I  L  T  C  P  F  E  P  P  K  P  K  T  K  H  K  L  D

GTCATGTCTGTGGAGGACTACAAAGCCCTGCAGAAGTACGAAAAGGAGAAGTTTGAAGAGATGATTAAGCAGATTAAAGAAACTGGTGCT
V  M  S  V  E  D  Y  K  A  L  Q  K  Y  E  K  E  K  F  E  E  M  I  K  Q  I  K  E  T  G  A

AACCTAGCTATTTGCCAGTGGGGCTTTGACGATGAAGCCAATCACTTACTTCTTCAGAACGGCCTGCCTGCAGTCCGCTGGGTAGGGGGA
N  L  A  I  C  Q  W  G  F  D  D  E  A  N  H  L  L  L  Q  N  G  L  P  A  V  R  W  V  G  G

CCTGAGATTGAGCTGATCGCCATTGCAACAGGAGGACGGATTGTCCCACGGTTCTCAGAGCTCACCTCTGAGAAGCTGGGCTTTGCTGGT
P  E  I  E  L  I  A  I  A  T  G  G  R  I  V  P  R  F  S  E  L  T  S  E  K  L  G  F  A  G

GTGGTGCAGGAGATCTCCTTTGGCACTACAAAAGACAAAATGCTGGTTATCGAGAAGTGTAAGAACTCTAGAGCTGTGACCATTTTCATC
V  V  Q  E  I  S  F  G  T  T  K  D  K  M  L  V  I  E  K  C  K  N  S  R  A  V  T  I  F  I

AGAGGAGGAAACAAGATGATCATAGAAGAAGCAAAACGATCTCTCCATGATGCCCTGTGTGTCATCCGGAACCTCATCCGTGACAACCGT
R  G  G  N  K  M  I  I  E  E  A  K  R  S  L  H  D  A  L  C  V  I  R  N  L  I  R  D  N  R

GTTGTGTATGGAGGAGGGGCAGCCGAAATATCCTGCGCCCTGGCAGTCAGCCAAGAGGCAGACAAGTGCCCAACTTTGGAACAGTATGCC
V  V  Y  G  G  G  A  A  E  I  S  C  A  L  A  V  S  Q  E  A  D  K  C  P  T  L  E  Q  Y  A

ATGAGAGCTTTTGCAGATGCCTTGGAGGTCATCCCCATGGCCCTTTCAGAAAATAGTGGCATGAATCCCATTCAGACCATGACTGAAGTT
M  R  A  F  A  D  A  L  E  V  I  P  M  A  L  S  E  N  S  G  M  N  P  I  Q  T  M  T  E  V

CGAGCCAGACAGGTGAAGGAGTCTAACCCTGCCCTGGGGATTGACTGTTTGCACAAGGGCAGTAACGATATGCAGTATCAGCATGTCATA
R  A  R  Q  V  K  E  S  N  P  A  L  G  I  D  C  L  H  K  G  S  N  D  M  Q  Y  Q  H  V  I

GAAACCTTGATTGGCAAAAAGCAGCAGATCTCTCTTGCCACCCAGATGGTTAGGATGATTCTGAAGATTGATGACATCCGTAAGCCTGGA
E  T  L  I  G  K  K  Q  Q  I  S  L  A  T  Q  M  V  R  M  I  L  K  I  D  D  I  R  K  P  G

GAATCTGAAGAATAAACTGTACCATTTACCACTGTGACTAAATAAAGGGTGTGTCTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
E  S  E  E  .
```

Fig.8f.

(Cctz)

pTzeta12 cDNA sequence

```
GAAGACCCCGCAGAGAGCACGTTGTTCTCGGCCTCTCCCGGAGCTAGGCCAGCCATGGCGGCGGTAAAGACCCTAAATCCGAAGGCCGAG
                                                    M  A  A  V  K  T  L  N  P  K  A  E

GTGGCCCGGGCCCAGGCAGCGCTGGCGGTGAACATCAGCGCGGCTCGGGGCCTGCAGGATGTTCTGAGGACCAACTTGGGGCCTAAGGGC
V  A  R  A  Q  A  A  L  A  V  N  I  S  A  A  R  G  L  Q  D  V  L  R  T  N  L  G  P  K  G

ACCATGAAGATGCTTGTATCTGGTGCTGGAGACATCAAACTTACTAAAGATGGCAATGTGCTGCTTCATGAAATGCAAATTCAACACCCA
T  M  K  M  L  V  S  G  A  G  D  I  K  L  T  K  D  G  N  V  L  L  H  E  M  Q  I  Q  H  P

ACAGCCTCTTTGATAGCAAAAGTGGCTACAGCCCAGGATGACATAACTGGCGATGGCACTACATCCAATGTCCTCATCATCGGGGAGCTG
T  A  S  L  I  A  K  V  A  T  A  Q  D  D  I  T  G  D  G  T  T  S  N  V  L  I  I  G  E  L

CTCAAACAGGCGGACCTGTACATTTCTGAAGGTCTTCACCCAAGAATAATAACTGAAGGTTTTGAAGCGGCAAAAGAAAAGGCACTCCAA
L  K  Q  A  D  L  Y  I  S  E  G  L  H  P  R  I  I  T  E  G  F  E  A  A  K  E  K  A  L  Q

TTTCTGGAACAAGTCAAAGTAAGCAAAGAGATGGACAGAGAAACACTCATCGATGTGGCCAGGACATCTCTGCGGACTAAAGTTCATGCT
F  L  E  Q  V  K  V  S  K  E  M  D  R  E  T  L  I  D  V  A  R  T  S  L  R  T  K  V  H  A

GAACTTGCAGATGTCTTGACAGAGGCTGTAGTGGACTCCATCTTGGCCATTAGGAAAAAGGACGAGCCCATTGACCTCTTCATGGTTGAG
E  L  A  D  V  L  T  E  A  V  V  D  S  I  L  A  I  R  K  K  D  E  P  I  D  L  F  M  V  E

ATCATGGAGATGAAGCATAAATCTGAGACAGATACAAGCTTAATCAGAGGGCTTGTTTTGGATCATGGAGCTCGGCATCCTGATATGAAG
I  M  E  M  K  H  K  S  E  T  D  T  S  L  I  R  G  L  V  L  D  H  G  A  R  H  P  D  M  K

AAGAGAGTGGAAAATGCCTACATCCTCACGTGCAACGTGTCCTTAGAGTATGAGAAAACAGAAGTGAATTCTGGGTTTTTTTACAAGAGT
K  R  V  E  N  A  Y  I  L  T  C  N  V  S  L  E  Y  E  K  T  E  V  N  S  G  F  F  Y  K  S

GCAGAAGAGAGAGAAAAACTAGTAAAGGCTGAAAGAAAATTCATTGAAGATAGAGTTAAAAAAATCATAGAGCTGAAAAAGAAAGTCTGT
A  E  E  R  E  K  L  V  K  A  E  R  K  F  I  E  D  R  V  K  K  I  I  E  L  K  K  K  V  C

GGTGACTCAGATAAAGGATTTGTCGTTATTAATCAAAAGGGGATTGACCCCTTTTCCTTAGATGCCCTTGCGAAAGAAGGGATCGTAGCT
G  D  S  D  K  G  F  V  V  I  N  Q  K  G  I  D  P  F  S  L  D  A  L  A  K  E  G  I  V  A

CTGCGCAGAGCCAAGAGGAGAAACATGGAGAGGCTGACACTTGCTTGTGGTGGGATAGCTCTGAATTCCTTTGATGACCTGAATCCTGAC
L  R  R  A  K  R  R  N  M  E  R  L  T  L  A  C  G  G  I  A  L  N  S  F  D  D  L  N  P  D

TGTTTGGGACATGCAGGGCTTGTCTATGAGTATACACTGGGTGAGGAGAAGTTCACCTTTATTGAGAAGTGTAACAATCCCCGTTCTGTC
C  L  G  H  A  G  L  V  Y  E  Y  T  L  G  E  E  K  F  T  F  I  E  K  C  N  N  P  R  S  V

ACTTTACTGGTTAAAGGACCAAATAAGCACACACTGACTCAAATCAAGGATGCAATAAGAGATGGCTTGAGGGCTGTCAAAAATGCTATT
T  L  L  V  K  G  P  N  K  H  T  L  T  Q  I  K  D  A  I  R  D  G  L  R  A  V  K  N  A  I

GATGATGGCTGTGTTGTCCCAGGTGCGGGTGCAGTAGAAGTGGCACTGGCAGAAGCTCTGATTAAATACAAGCCCAGTGTGAAGGGCAGG
D  D  G  C  V  V  P  G  A  G  A  V  E  V  A  L  A  E  A  L  I  K  Y  K  P  S  V  K  G  R

GCGCAGCTTGGAGTCCAGGCATTTGCAGATGCCTTGCTCATCATTCCCAAGGTTCTTGCGCAAAACTCTGGTTTTGACCTTCAGGAAACA
A  Q  L  G  V  Q  A  F  A  D  A  L  L  I  I  P  K  V  L  A  Q  N  S  G  F  D  L  Q  E  T

TTAGTTAAAGTTCAAGCTGAACATTCAGAATCGGGCCAGCTCGTAGGTGTGGATCTGAGCACAGGTGAGCCGATGGTGGCCGCAGAGATG
L  V  K  V  Q  A  E  H  S  E  S  G  Q  L  V  G  V  D  L  S  T  G  E  P  M  V  A  A  E  M

GGTGTGTGGGATAACTACTGTGTGAAGAAGCAGCTGCTACACTCCTGTACTGTGATCGCCACCAACATTCTCCTGGTCGACGAGATCATG
G  V  W  D  N  Y  C  V  K  K  Q  L  L  H  S  C  T  V  I  A  T  N  I  L  L  V  D  E  I  M

CGAGCTGGGATGTCCTCTCTGAAGGGTTGAGGCCTGCCTGTGATACTACAGGATGTTGGGGGGAATGGTTATTTTTGTCCAAGCTTCAAG
R  A  G  M  S  S  L  K  G  .

TGATTTGGAAAAAAATTTTCTCTTCCTGATTGGAGAAAAGAAACGGGACATTTGACACCTATTCAAATTATACTGTAAAATTTTATTTTA

TTTTTGCCTTGAGTATCTGAAGACACTCAAAGCAGCTCTTTTTCAACCCACTGAACAAGATGTTTTAGCTACACCGATACAAAAATTACA

TAATAAGATAAGCATGTTGTCTACCCTTGTTCCATAAGTGTTCTTTGAAAGTTTGTAATGGTTTTCTCCTAAATAAGGCATGGTGACACA

TGCCTGTAAGCCTAGCCCTTTGGAAATAGTCCGGAATTTCTATGCCAACTCAGGCTACAGGAGACCCCAGGTCGAAAGAATAATTTGTTG

TGGATGTATTTGAAATTATCCAGCCAACTCCCTGTTAAACATGTAAGATCCTTGCCAGTGTAAAACACATCTGGGTAATTTATGGGTTGC

ATAATGTCTAATAAATACTTAAAAGAGTGAAAAAAAAAAAAAAAA
```

Fig.8g.

(Ccth)

pCBL80 cDNA sequence

```
AGATGATGCCCACACCAGTTATCCTGTTGAAAGAGGGTACTGATAGCTCCCAGGGCATCCCTCAGCTCGTGAGTAACATCAGTGCCTGC
  M  M  P  T  P  V  I  L  L  K  E  G  T  D  S  S  Q  G  I  P  Q  L  V  S  N  I  S  A  C

CAAGTGATTGCTGAGGCTGTAAGAACCACCCTGGGTCCCCGTGGCATGGACAAACTTATTGTGGATGGCCGAGGCAAAGCAACAATATCT
Q  V  I  A  E  A  V  R  T  T  L  G  P  R  G  M  D  K  L  I  V  D  G  R  G  K  A  T  I  S

AATGATGGGGCCACAATTCTGAAACTCCTCGATGTTGTCCATCCTGCAGCAAAGACTTTAGTGGACATAGCCAAGTCCCAGGATGCTGAG
N  D  G  A  T  I  L  K  L  L  D  V  V  H  P  A  A  K  T  L  V  D  I  A  K  S  Q  D  A  E

GTTGGTGATGGCACCACCTCAGTGACCCTGCTGGCTGCGGAGTTTCTGAAGCAGGTGAAGCCCTACGTGGAAGAAGGTTTACACCCTCAG
V  G  D  G  T  T  S  V  T  L  L  A  A  E  F  L  K  Q  V  K  P  Y  V  E  E  G  L  H  P  Q

ATCATCATCCGAGCTTTCCGCACAGCCACCCAATTGGCTGTTAACAAAATCAAAGAGATAGCTGTGACTGTGAAGAAGCAAGATAAAGTA
I  I  I  R  A  F  R  T  A  T  Q  L  A  V  N  K  I  K  E  I  A  V  T  V  K  K  Q  D  K  V

GAGCAGAGGAAGATGCTGGAGAAGTGTGCGATGACAGCCCTGAGCTCCAAGCTGATCTCCCAGCAGAAGGTCTTCTTCGCCAAGATGGTG
E  Q  R  K  M  L  E  K  C  A  M  T  A  L  S  S  K  L  I  S  Q  Q  K  V  F  F  A  K  M  V

GTTGATGCCGTGATGATGCTTGACGAGCTGCTGCAGCTTAAAATGATTGGCATCAAGAAGGTGCAGGGGGGAGCCCTGGAGGAGTCTCAG
V  D  A  V  M  M  L  D  E  L  L  Q  L  K  M  I  G  I  K  K  V  Q  G  G  A  L  E  E  S  Q

CTAGTTGCTGGTGTTGCGTTCAAGAAGACTTTCTCTTATGCTGGGTTTGAAATGCAGCCCAAGAAGTATAAGAACCCCAAGATTGCCCTC
L  V  A  G  V  A  F  K  K  T  F  S  Y  A  G  F  E  M  Q  P  K  K  Y  K  N  P  K  I  A  L

TTAAATGTTGAGCTTGAGCTGAAAGCAGAGAAAGATAATGCTGAAATCAGAGTCCACACAGTGGAGGATTACCAGGCAATTGTTGATGCC
L  N  V  E  L  E  L  K  A  E  K  D  N  A  E  I  R  V  H  T  V  E  D  Y  Q  A  I  V  D  A

GAGTGGAATATTCTCTATGACAAGTTAGAGAAGATCCATCAGTCTGGAGCCAAAGTCATCTTGTCTAAACTCCCTATTGGGGATGTGGCC
E  W  N  I  L  Y  D  K  L  E  K  I  H  Q  S  G  A  K  V  I  L  S  K  L  P  I  G  D  V  A

ACCCAGTACTTTGCTGATAGGGACATGTTCTGTGCTGGCCGAGTGCCTGAGGAGGATCTGAAGAGGACGATGATGGCTTGTGGAGGCTCA
T  Q  Y  F  A  D  R  D  M  F  C  A  G  R  V  P  E  E  D  L  K  R  T  M  M  A  C  G  G  S

ATCCAGACCAGTGTGAATGCTCTGGTTCCAGATGTGCTGGGCCACTGCCAAGTGTTTGAAGAGACCCAAATTGGAGGAGAGAGGTACAAT
I  Q  T  S  V  N  A  L  V  P  D  V  L  G  H  C  Q  V  F  E  E  T  Q  I  G  G  E  R  Y  N

TTCTTCACTGGCTGCCCTAAGGCCAAGACATGTACCATCATCCTCCGTGGTGGCGCTGAGCAGTTTATGGAAGAGACAGAGAGGTCCCTA
F  F  T  G  C  P  K  A  K  T  C  T  I  I  L  R  G  G  A  E  Q  F  M  E  E  T  E  R  S  L

CATGATGCCATCATGATTGTGAGGAGGGCCATCAAGAATGACTCTGTGGTGGCTGGTGGTGGAGCCATCGAGATGGAACTTTCCAAATAC
H  D  A  I  M  I  V  R  R  A  I  K  N  D  S  V  V  A  G  G  G  A  I  E  M  E  L  S  K  Y

CTGCGGGATTACTCGAGGACCATTCCTGGGAAGCAGCAGCTGTTGATTGGGGCATATGCCAAGGCCCTGGAGATTATTCCACGACAGCTA
L  R  D  Y  S  R  T  I  P  G  K  Q  Q  L  L  I  G  A  Y  A  K  A  L  E  I  I  P  R  Q  L

TGTGACAACGCTGGCTTTGATGCCACAAACATCCTCAACAAGCTGCGGGCTCGACACGCACAGGGAGGTATGTGGTATGGGGTGGACATC
C  D  N  A  G  F  D  A  T  N  I  L  N  K  L  R  A  R  H  A  Q  G  G  M  W  Y  G  V  D  I

AACAATGAGAACATCGCCGACAACTTCCAGGCATTTGTGTGGGAGCCAGCCATGGTGCGCATCAACGCTCTGACAGCAGCTTCTGAGGCT
N  N  E  N  I  A  D  N  F  Q  A  F  V  W  E  P  A  M  V  R  I  N  A  L  T  A  A  S  E  A

GCATGCCTTATTGTGTCCGTGGATGAGACTATCAAGAACCCCCGCTCCACTGTGGATCCTCCAGCTCCATCAGCTGGCCGTGGCAGAGGC
A  C  L  I  V  S  V  D  E  T  I  K  N  P  R  S  T  V  D  P  P  A  P  S  A  G  R  G  R  G

CAAGCCCGCTTCCACTGAGAGGCGAGGCGGTCTGCACCTCCTTGTGAGGTGAGGGGGTGGATGAGAAGATGGTTGCTGGTCTGCTGGGTT
Q  A  R  F  H  .

CTCACTGAGGTTATTTAAATAAAGC
```

Fig.8h.
(Cctq)

pTtheta1 cDNA sequence

```
                                                                  M
CCCATTCCCAGCAGGCAGTGCGCGGCGCCCCTTGCTTCCTCGGCGTCTCGGAGCGAGCGCGTGAGCGGTGGACTGCGGCCGAGCCATG
                                                                  M
GCGCTTCACGTCCCCAAGGCCCCGGGCTTTGCCCAGATGCTCAAGGATGGAGCCAAACATTTCTCGGGATTAGAAGAGGCTGTGTATAGA
A  L  H  V  P  K  A  P  G  F  A  Q  M  L  K  D  G  A  K  H  F  S  G  L  E  E  A  V  Y  R

AATATACAGGCCTGCAAGGAGCTTGCTCAGACTACTCGCACAGCGTACGGACCAAATGGAATGAATAAAATGGTCATCAATCGCCTGGAG
N  I  Q  A  C  K  E  L  A  Q  T  T  R  T  A  Y  G  P  N  G  M  N  K  M  V  I  N  R  L  E

AAGTTGTTTGTGACAAACGATGCAGCGACTATTTTAAGAGAGCTAGAAGTGCAGCATCCTGCTGCAAAGATGATAGTGATGGCCTCTCAT
K  L  F  V  T  N  D  A  A  T  I  L  R  E  L  E  V  Q  H  P  A  A  K  M  I  V  M  A  S  H

ATGCAAGAACAAGAGGTTGGTGATGGCACAAACTTCGTTCTGGTGTTCGCTGGGGCTCTTCTGGAACTAGCTGAAGAACTTCTGAGGATT
M  Q  E  Q  E  V  G  D  G  T  N  F  V  L  V  F  A  G  A  L  L  E  L  A  E  E  L  L  R  I

GGCCTGTCAGTATCAGAGGTAATATCGGGGTATGAAATAGCTTGCAAAAAAGCTCATGAGATCCTTCCTGAGTTGGTATGTTGTTCTGCC
G  L  S  V  S  E  V  I  S  G  Y  E  I  A  C  K  K  A  H  E  I  L  P  E  L  V  C  C  S  A

AAAAACCTTCGAGATGTTGATGAAGTGTCCTCTCTGCTTCGAACCTCCATCATGAGTAAGCAGTACGGCAGCGAGACGTTTTTGGCCAAG
K  N  L  R  D  V  D  E  V  S  S  L  L  R  T  S  I  M  S  K  Q  Y  G  S  E  T  F  L  A  K

CTTATTGCTCAGGCTTGTGTTTCTATCTTTCCTGATTCTGGCAATTTCAATGTTGATAACATCAGAGTATGTAAGATTCTGGGCTCTGGT
L  I  A  Q  A  C  V  S  I  F  P  D  S  G  N  F  N  V  D  N  I  R  V  C  K  I  L  G  S  G

ATTTATTCATCCTCAGTATTACATGGCATGGTTTTTAAGAAGGAAACTGAAGGTGATGTGACATCTGTCAAAGATGCAAAGATAGCTGTG
I  Y  S  S  S  V  L  H  G  M  V  F  K  K  E  T  E  G  D  V  T  S  V  K  D  A  K  I  A  V

TACTCTTGTCCGTTTGATGGCATGATAACAGAGACAAAGGGGACCGTGCTGATTAAGACTGCCGAGGAGCTAATGAACTTCAGTAAGGGA
Y  S  C  P  F  D  G  M  I  T  E  T  K  G  T  V  L  I  K  T  A  E  E  L  M  N  F  S  K  G

GAGGAGAACCTCATGGATGCTCAGGTGAAGGCCATTGCAGGCACTGGTGCAAATGTCATAGTAACAGGCGGCAAAGTGGCGGACATAGCT
E  E  N  L  M  D  A  Q  V  K  A  I  A  G  T  G  A  N  V  I  V  T  G  G  K  V  A  D  I  A

CTTCATTATGCTAACAAGTACAATATCATGTTGGTGAGACTGAACTCAAAGTGGGATCTCAGACGACTCTGTAAAACAGTTGGTGCCACA
L  H  Y  A  N  K  Y  N  I  M  L  V  R  L  N  S  K  W  D  L  R  R  L  C  K  T  V  G  A  T

GCTCTTCCAAAATTGACTCCTCCCGTCCAAGAAGAAATGGGACATTGTGACAGTGTTTACCTCTCAGAAGTTGGAGATACACAAGTGGTT
A  L  P  K  L  T  P  P  V  Q  E  E  M  G  H  C  D  S  V  Y  L  S  E  V  G  D  T  Q  V  V

GTTTTTAAGCATGAAAAAGAAGATGGTGCCATTTCTACTATAGTTCTTCGAGGTTCTACAGACAATCTGATGGATGACATAGAAAGGGCA
V  F  K  H  E  K  E  D  G  A  I  S  T  I  V  L  R  G  S  T  D  N  L  M  D  D  I  E  R  A

GTAGATGATGGAGTTAATACTTTCAAAGTTCTCACAAGGGATAAGCGTCTTGTACCTGGAGGTGGAGCTACCGAAATTGAATTGGCTAAA
V  D  D  G  V  N  T  F  K  V  L  T  R  D  K  R  L  V  P  G  G  G  A  T  E  I  E  L  A  K

CAAATCACATCATATGGAGAGACGTGTCCTGGGCTTGAACAGTATGCTATTAAGAAGTTTGCTGAAGCGTTTGAAGCGATTCCACGGGCA
Q  I  T  S  Y  G  E  T  C  P  G  L  E  Q  Y  A  I  K  K  F  A  E  A  F  E  A  I  P  R  A

CTGGCAGAAAATTCTGGCGTGAAGGCCAATGAAGTTATCTCTAAACTTTATTCCGTACACCAAGAAGGAAACAAAAATGTGGGGTTGGAT
L  A  E  N  S  G  V  K  A  N  E  V  I  S  K  L  Y  S  V  H  Q  E  G  N  K  N  V  G  L  D

ATCGAGGCTGAAGTCCCTGCTGTAAAGGATATGTTAGAAGCCAGTATTTTAGATACTTACTTGGGAAAATACTGGGCTATTAAACTGGCC
I  E  A  E  V  P  A  V  K  D  M  L  E  A  S  I  L  D  T  Y  L  G  K  Y  W  A  I  K  L  A

ACTAATGCTGCAGTCACTGTACTAAGAGTGGATCAGATCATCATGGCAAAACCAGCTGGTGGGCCCAAACCTCCAAGTGGGAAGAAGGAC
T  N  A  A  V  T  V  L  R  V  D  Q  I  I  M  A  K  P  A  G  G  P  K  P  P  S  G  K  K  D

TGGGATGACGACCAGAATGACTGAGGAACTTGTCATAGGTTAAGAGTTGTGTTTGTAGAGTAGAACTTGCCCAGTGTTTTATTTTTCTTA
W  D  D  D  Q  N  D  .

TTTGTGTTTGTACTCTGCTGGGTGGTACAATAAATGTGTTACTGTTAAAAAAAAA
```

Fig.9.

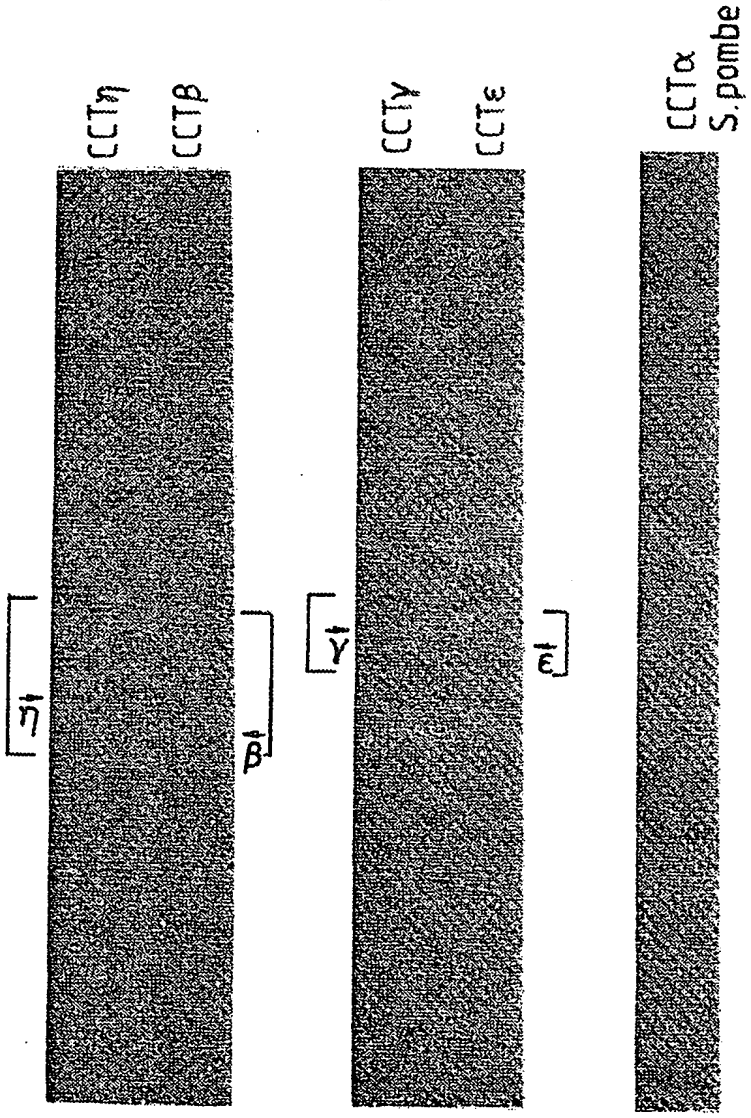


Fig.10.

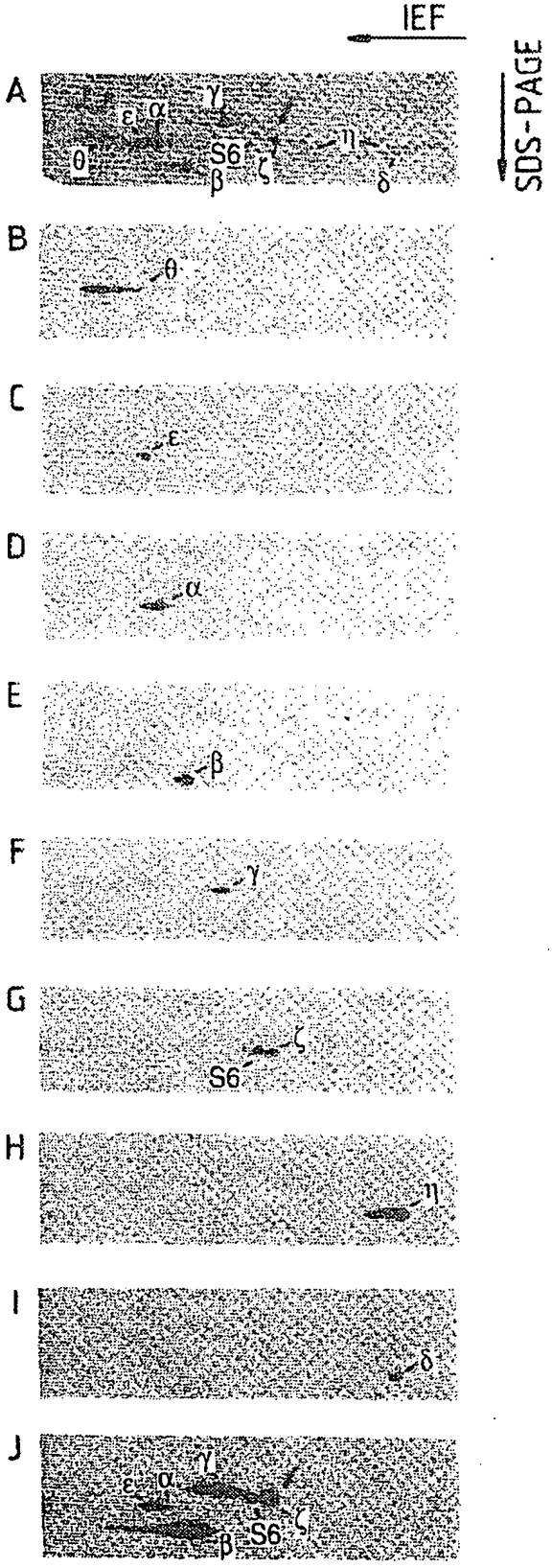


Fig.11a.

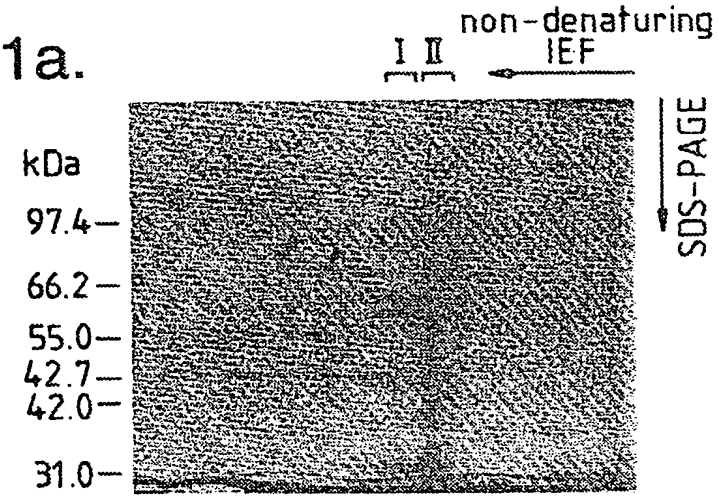


Fig.11b.

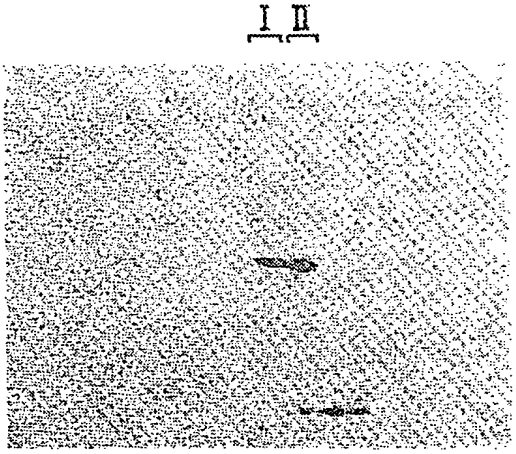


Fig.11c.

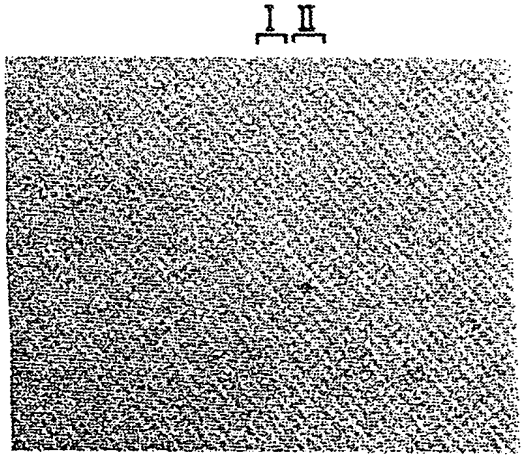


Fig.12.

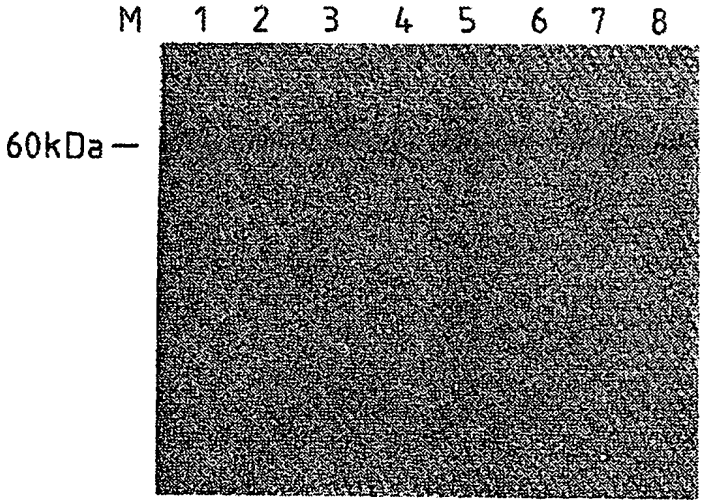


Fig.13a.

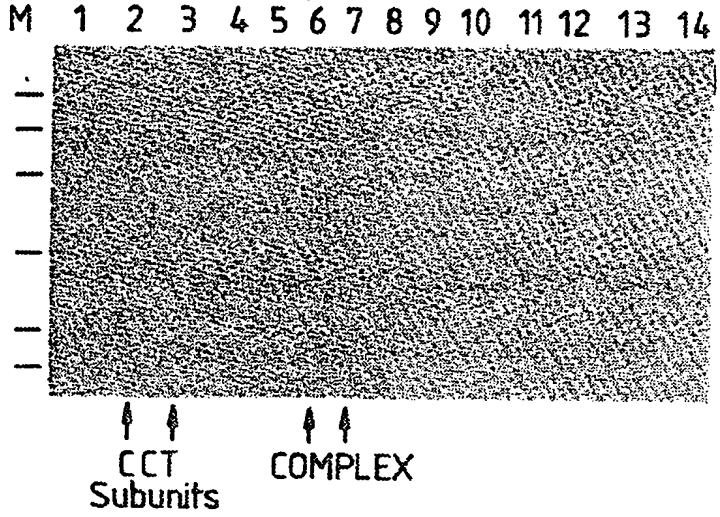


Fig.13b.